Active Learning of Conditional Mean Embeddings via Bayesian Optimisation

Sayak Ray Chowdhury[*]
Indian Institute of Science
sayak@iisc.ac.in

Rafael Oliveira[*]
The University of Sydney
rafael.oliveira@sydney.edu.au

Fabio Ramos
The University of Sydney & NVIDIA
fabio.ramos@sydney.edu.au

Abstract

We consider the problem of sequentially optimising the conditional expectation of an objective function, with both the conditional distribution and the objective function assumed to be fixed but unknown. Assuming that the objective function belongs to a reproducing kernel Hilbert space (RKHS), we provide a novel upper confidence bound (UCB) based algorithm CME-UCB via estimation of the *conditional mean embeddings* (CME), and derive its regret bound. Along the way, we derive novel approximation guarantees for the CME estimates. Finally, experiments are carried out in a synthetic example and in a likelihood-free inference application that highlight the useful insights of the proposed method.

1 INTRODUCTION

A large class of problems in machine learning and statistics involve the optimisation of an expected objective conditioned on control variables. For instance, one may want to minimise expected risks (Beyer and Sendhoff, 2007) or maximise the expected rewards in reinforcement learning (Deisenroth et al., 2011). Alternatively, the involved conditional distributions themselves might be of interest, as in likelihood-free inference (Gutmann and Corander, 2016; Papamakarios et al., 2019). These are hard-to-model stochastic processes which are unsuitable for standard optimisation algorithms.

In recent years the representation of conditional distributions $p(\mathbf{x}|\mathbf{u})$ as elements of a reproducing kernel Hilbert space (RKHS), known as *conditional mean embeddings*, has become increasingly popular (Song et al., 2009). In this formulation the conditional expectation of any function f in the RKHS becomes a linear operation, via the RKHS inner product with the appropriate distribution embedding. Conditional mean embeddings have been successfully applied to many machine learning tasks such as hidden Markov models (Song et al., 2010a), non-parametric graphical models (Song et al., 2010b), modelling transition dynamics in MDPs (Grünewälder et al., 2012b), subspace selection (Fukumizu et al., 2009) and conditional independence testing (Fukumizu et al., 2008). We refer the interested reader to the monograph by Muandet et al. (2016) for a review.

The main motivation behind conditional mean embeddings has been to generalise the notion of conditional expectation to Hilbert spaces. Its foremost advantage is that one can directly compute conditional expectations based on the observed data. The alternative approach of learning a conditional density estimate as an intermediate step scales poorly with the dimension of the underlying space (Grünewälder et al., 2012b). Additionally, conditional mean embeddings can be characterised as the solution of a Tikhonov regularized vector-valued regression problem with the square loss (Grünewälder et al., 2012a). Convergence of conditional mean embeddings in RKHS norm has been established under *independent and identically distributed* (i.i.d.) samples (Song et al., 2010b; Grünewälder et al., 2012a), which essentially shows that the estimated embeddings are consistent under certain smoothness assumptions. However, in an active or sequential learning environment like the one considered in this work, one collects data based on past observations, and hence existing bounds fail to remain useful.

Against this backdrop, we revisit the problem of sequentially maximising the conditional expectation of a function in an RKHS via estimating the conditional mean embedding. We make the following contributions:

- Under non-i.i.d. samples, we derive a concentration bound on conditional mean embeddings and their

[*]Equal contribution

Proceedings of the 36[th] Conference on Uncertainty in Artificial Intelligence (UAI), PMLR volume 124, 2020.

estimators in RKHS norm as a function of the uncertainties around these estimates (Theorem 1). This bound not only serves as a key tool in designing our algorithm but also is of independent interest.

- We develop an algorithm, namely *Conditional Mean Embeddings Upper Confidence Bound* (CME-UCB), for maximising the conditional expectation of a function (Algorithm 1) and derive a high-probability regret bound under RKHS regularity assumption (Theorem 4).
- Finally, we experimentally verify our results in a synthetic toy example and also in a likelihood-free inference application, for which our algorithm is seen to perform favourably.

2 RELATED WORK

Black-box optimisation of an unknown function with expensive, noisy queries is a generic problem arising in domains such as hyper-parameter tuning for complex machine learning models (Snoek et al., 2012), policy search (Wilson et al., 2014), environmental monitoring (Marchant and Ramos, 2012), experimental design etc. Bayesian optimization (BO), the most popular approach towards solving this problem, starts with a prior distribution over a function class, typically a Gaussian process (GP) (Rasmussen and Williams, 2006), uses function evaluations to compute the posterior distribution over functions, and chooses the next function evaluation adaptively towards reaching the optimum. Perhaps the most prominent algorithm with rigorous theoretical guarantees in this regard is the Gaussian process upper confidence bound (GP-UCB) algorithm (Srinivas et al., 2010). Recently, Oliveira et al. (2019) considers the BO problem under input noise, which has applications in certain areas of robotics and process control, and design an algorithm via a GP model that takes conditional distributions as inputs. Their method, however, does not attempt to learn the distribution model.

Other approaches to learning kernel embeddings focus on learning the hyper-parameters of the kernel, which are critical in certain applications. Flaxman et al. (2016) propose a Bayesian approach to learn kernel embedding hyper-parameters by means of the marginal likelihood of a GP, which is placed as a prior over the true embedding. Alternatively, Hsu and Ramos (2019) propose a closed-form approach to estimate probability density functions in a Bayesian probabilistic model. The latter allows them to learn hyper-parameters by maximising the evidence of the data. Buathong et al. (2019) consider maximising the conditional expectation of an unknown function, with the expectation being taken over the uniform distribution on a given set. Despite that, all of these methods still use i.i.d. data to form the distribution embeddings. Our concern in this paper is in deriving a more efficient data collection process which can improve the estimates of both the conditional distribution and an underlying objective function. Lastly, Vien and Toussaint (2018) consider optimising a function defined over an RKHS, while we focus on optimising the conditional expectation of a function with a goal to find the optimal control.

3 PROBLEM STATEMENT

We consider the optimisation problem of maximising a function $f : \mathcal{X} \to \mathbb{R}$ over a given state space $\mathcal{X} \subset \mathbb{R}^D$. However, we assume no direct control over the space $\mathcal{X}$. Instead, we can only set control variables $\mathbf{u} \in \mathcal{U}$ in a given control space $\mathcal{U} \subset \mathbb{R}^d$. The application of a control $\mathbf{u}$ results in a state $\mathbf{x}$ distributed according to $\mathbf{x}|\mathbf{u} \sim P_{\mathbf{u}}$. In addition, both f and the mapping $\mathbf{u} \mapsto P_{\mathbf{u}}$ are unknown. The algorithm can select up to n controls to find:

$$\mathbf{u}^\star \in \operatorname*{argmax}_{\mathbf{u}\in\mathcal{U}} \mathbb{E}[f(\mathbf{x})|\mathbf{u}] \,. \tag{1}$$

The query point $\mathbf{u}_t$ at time t is chosen causally depending upon the past observations $\mathcal{D}_{t-1} = \{(\mathbf{u}_i, \mathbf{x}_i, y_i)\}_{i=1}^{t-1}$. For each $\mathbf{u}_t$, the algorithm is provided with observations $y_t = f(\mathbf{x}_t) + \zeta_t$, where $\mathbf{x}_t|\mathbf{u}_t \sim P_{\mathbf{u}_t}$. We assume that ζ_t is zero-mean conditionally σ_ζ-sub-Gaussian observation noise, for some $\sigma_\zeta \geq 0$. More precisely,

$$\forall\gamma \in \mathbb{R}, \quad \mathbb{E}[\exp(\gamma\zeta_t)\,|\mathcal{H}_{t-1}] \leq \exp\left(\gamma^2\sigma_\zeta^2/2\right) \text{ (a.s.)}, \tag{2}$$

where $\mathcal{H}_{t-1}$ is the σ-algebra generated by the random variables $\{(\mathbf{x}_i, y_i)\}_{i=1}^{t-1}$ and $\mathbf{x}_t$ and the expectation holds in the almost surely (a.s.) sense. One common metric to evaluate the performance of any sequential algorithm is the cumulative *regret*, defined in our context as:

$$R_n = \sum\nolimits_{t=1}^{n} r_t = \sum\nolimits_{t=1}^{n} \mathbb{E}[f(\mathbf{x})|\mathbf{u}^\star] - \mathbb{E}[f(\mathbf{x})|\mathbf{u}_t] \,.$$

A sublinear growth of R_n with n implies the time-average regret $R_n/n \to 0$ as $n \to \infty$. The latter indicates that the algorithm is able to get arbitrarily close to the optimum over time, since $\min_{t\leq n} r_t \leq R_n/n$.

Regularity assumptions: We assume $f : \mathcal{X} \to \mathbb{R}$ to be an element of $\mathcal{H}_k$, which is a reproducing kernel Hilbert space (RKHS) (Schölkopf and Smola, 2002). For a given positive-definite kernel $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, a RKHS $\mathcal{H}_k$ is a Hilbert space of functions with feature map $\phi_k : \mathcal{X} \to \mathcal{H}_k$, inner product $\langle\cdot,\cdot\rangle_k$ and norm $\|\cdot\|_k = \sqrt{\langle\cdot,\cdot\rangle_k}$ such that $f(\mathbf{x}) = \langle f, k(\cdot,\mathbf{x})\rangle_k$ and $k(\mathbf{x},\mathbf{x}') = \langle\phi_k(\mathbf{x}), \phi_k(\mathbf{x}')\rangle_k$ for any $f \in \mathcal{H}_k$ and any $\mathbf{x}, \mathbf{x}' \in \mathcal{X}$. We assume k is continuous and bounded on

$\mathcal{X} \times \mathcal{X}$, with $k(\mathbf{x}, \mathbf{x}) \leq 1, \forall \mathbf{x} \in \mathcal{X}$, and that $\|f\|_k \leq b$ for the objective function in Equation 1, where $b > 0$ is known. Boundedness of k along the diagonal holds for any stationary kernel, i.e., where $k(\mathbf{x}, \mathbf{x}') = k(\mathbf{x} - \mathbf{x}')$, e.g., the *squared exponential* kernel and the *Matérn* kernel (Rasmussen and Williams, 2006).

Distribution assumptions: Let $\mathcal{P}$ denote the set of all probability measures on $\mathcal{X}$. With $f \in \mathcal{H}_k$, we can define the map:

$$\begin{aligned} \psi : \mathcal{P} &\to \mathcal{H}_k \\ P &\mapsto \int_{\mathcal{X}} k(\cdot, \mathbf{x}) \, \mathrm{d}P(\mathbf{x}) \; . \end{aligned} \tag{3}$$

For any $\mathcal{X}$-valued random variable $\mathbf{x}$ distributed according to $P \in \mathcal{P}$, we then have that:

$$\mathbb{E}_P[f] := \mathbb{E}[f(\mathbf{x})] = \langle f, \psi_P \rangle_k, \quad \forall f \in \mathcal{H}_k \; , \tag{4}$$

where $\psi_P := \psi(P)$. If the kernel k is characteristic, such as radial kernels (Sriperumbudur et al., 2011), ψ is injective, defining a one-to-one relationship between measures in $\mathcal{P}$ and elements of $\mathcal{H}_k$. Therefore, ψ is referred to as the mean map, and ψ_P as the kernel mean embedding of P (Muandet et al., 2016). Now we define the map:

$$\begin{aligned} \vartheta : \mathcal{U} &\to \mathcal{H}_k \\ \mathbf{u} &\mapsto \psi_{P_\mathbf{u}} \; , \end{aligned} \tag{5}$$

where $P_\mathbf{u}$ represents a conditional probability distribution over $\mathcal{X}$ conditioned on $\mathbf{u} \in \mathcal{U}$. We then have that $\vartheta(\mathbf{u}) = \mathbb{E}[k(\cdot, \mathbf{x})|\mathbf{u}]$, and:

$$\forall f \in \mathcal{H}_k, \quad \mathbb{E}_{P_\mathbf{u}}[f] := \mathbb{E}[f(\mathbf{x})|\mathbf{u}] = \langle f, \vartheta(\mathbf{u}) \rangle_k \; . \tag{6}$$

We assume that the state distribution $P_\mathbf{u}$ for a given control $\mathbf{u}$ marginalises over all other variables that could affect the querying process, such as past states and effects from the environment where the agent is. In addition, we assume that the control space $\mathcal{U}$ is endowed with a positive definite kernel $c : \mathcal{U} \times \mathcal{U} \to \mathbb{R}$ with the associated RKHS $\mathcal{H}_c$ and the feature map $\phi_c : \mathcal{U} \to \mathcal{H}_c$. Then $\vartheta(\mathbf{u})$ traces out a set of mean embeddings in $\mathcal{H}_k$, one for each control $\mathbf{u}$, via a conditional embedding operator $\Theta : \mathcal{H}_c \to \mathcal{H}_k$, such that:

$$\vartheta(\mathbf{u}) = \Theta \phi_c(\mathbf{u}) \; . \tag{7}$$

The mapping $\Theta : \mathcal{H}_c \to \mathcal{H}_k$ can be seen as an element of the Hilbert space $\mathcal{H}_k \otimes \mathcal{H}_c$, whose reproducing kernel is $k(\mathbf{x}, \mathbf{x}')c(\mathbf{u}, \mathbf{u}')$.[1] We assume c is continuous and bounded, with $c(\mathbf{u}, \mathbf{u}) \leq 1, \forall \mathbf{u} \in \mathcal{U}$, and that Θ is bounded, i.e., $\|\Theta\|_{\mathrm{op}} \leq B$, where $B > 0$ is known.[2]

[1] This is a smoothness assumption on the conditional distribution $P_\mathbf{u}$ and is equivalent to assuming that $\mathbf{u} \mapsto \mathbb{E}_{P_\mathbf{u}}[f]$ is an element of $\mathcal{H}_c$ (Song et al., 2010a; Grünewälder et al., 2012b).

[2] $\|\cdot\|_{\mathrm{op}}$ denotes the operator norm: $\|\Theta\|_{\mathrm{op}} := \sup_{g \in \mathcal{H}_c : g \neq 0} \frac{\|\Theta g\|_k}{\|g\|_c}$.

4 ALGORITHM DESIGN

For a given $f \in \mathcal{H}_k$, the knowledge of the conditional distribution $P_\mathbf{u}$ would allow selecting points $\mathbf{u}_t$ based on the estimates for $\mathbb{E}_{P_\mathbf{u}}[f] = \langle f, \vartheta(\mathbf{u}) \rangle_k$. However, in general, the true mapping $\mathbf{u} \mapsto \vartheta(\mathbf{u})$ is unknown. Instead, we learn a model $\mathbf{u} \mapsto \hat{\vartheta}_t(\mathbf{u})$ based on the samples $\{\mathbf{u}_i, \mathbf{x}_i\}_{i=1}^t$. The conditional mean embedding operator can be estimated by solving the optimization problem:

$$\hat{\Theta}_t \in \operatorname*{argmin}_{\Theta : \mathcal{H}_c \to \mathcal{H}_k} \sum_{i=1}^{t} \|\phi_k(\mathbf{x}_i) - \Theta \phi_c(\mathbf{u}_i)\|_k^2 + \eta \|\Theta\|_{\mathrm{HS}}^2 \; , \tag{8}$$

where $\eta > 0$ is a regularising constant.[3] The solution of Equation 8 is given by:

$$\hat{\Theta}_t = \mathbf{\Phi}_k(\mathbf{X}_t) \mathbf{\Phi}_c(\mathbf{U}_t)^\mathsf{T} \left(\mathbf{\Phi}_c(\mathbf{U}_t) \mathbf{\Phi}_c(\mathbf{U}_t)^\mathsf{T} + \eta \mathbf{I} \right)^{-1} \; ,$$

where the columns of $\mathbf{\Phi}_k(\mathbf{X}_t)$ and $\mathbf{\Phi}_c(\mathbf{U}_t)$ contain the features $\phi_k(\mathbf{x}_i)$ and $\phi_c(\mathbf{u}_i)$, $1 \leq i \leq t$, respectively. Let $\mathbf{c}_t(\mathbf{u}) = \mathbf{\Phi}_c(\mathbf{U}_t)^\mathsf{T} \phi_c(\mathbf{u}) = [c(\mathbf{u}_1, \mathbf{u}), \ldots, c(\mathbf{u}_t, \mathbf{u})]^\mathsf{T}$ and $[\mathbf{C}_t]_{ij} = c(\mathbf{u}_i, \mathbf{u}_j)$, $1 \leq i, j \leq t$. Then the sample estimate of the conditional mean embedding becomes:

$$\hat{\vartheta}_t(\mathbf{u}) = \hat{\Theta}_t \phi_c(\mathbf{u}) = \mathbf{\Phi}_k(\mathbf{X}_t)(\mathbf{C}_t + \eta \mathbf{I})^{-1} \mathbf{c}_t(\mathbf{u}) \; . \tag{9}$$

Then, for a given $f \in \mathcal{H}_k$, $\mathbb{E}_{P_\mathbf{u}}[f]$ can be estimated by $\langle f, \hat{\vartheta}_t(\mathbf{u}) \rangle_k$. However, since the objective function f is also unknown, we need to estimate it as well. Given samples $\{\mathbf{x}_i, y_i\}_{i=1}^t$, by the representer theorem (Steinwart and Christmann, 2008), f can be estimated by:

$$\hat{\mu}_t = \mathbf{\Phi}_k(\mathbf{X}_t)(\mathbf{K}_t + \lambda \mathbf{I})^{-1} \mathbf{y}_t \; ,$$

where $\lambda > 0$ is a regularising constant, $[\mathbf{K}_t]_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$, $1 \leq i, j \leq t$ and $\mathbf{y}_t = [y_1, \ldots, y_t]^\mathsf{T}$. Let $\mathbf{k}_t(\mathbf{x}) = [k(\mathbf{x}_1, \mathbf{x}), \ldots, k(\mathbf{x}_t, \mathbf{x})]^\mathsf{T}$. By the reproducing property, we then have:

$$\hat{\mu}_t(\mathbf{x}) = \mathbf{k}_t(\mathbf{x})^\mathsf{T} (\mathbf{K}_t + \lambda \mathbf{I})^{-1} \mathbf{y}_t \; .$$

Now $\mathbb{E}_{P_\mathbf{u}}[f]$ can be estimated by:

$$\begin{aligned} \langle \hat{\mu}_t, \hat{\vartheta}_t(\mathbf{u}) \rangle_k &= \mathbf{c}_t(\mathbf{u})^\mathsf{T} (\mathbf{C}_t + \eta \mathbf{I})^{-1} \mathbf{K}_t (\mathbf{K}_t + \lambda \mathbf{I})^{-1} \mathbf{y}_t \\ &= \mathbf{c}_t(\mathbf{u})^\mathsf{T} (\mathbf{C}_t + \eta \mathbf{I})^{-1} \boldsymbol{\mu}_t \; , \end{aligned} \tag{10}$$

where we define:

$$\boldsymbol{\mu}_t = \mathbf{K}_t (\mathbf{K}_t + \lambda \mathbf{I})^{-1} \mathbf{y}_t = [\hat{\mu}_t(\mathbf{x}_1), \ldots, \hat{\mu}_t(\mathbf{x}_t)]^\mathsf{T} \; .$$

[3] $\|\cdot\|_{\mathrm{HS}}$ denotes the Hilbert-Schmidt norm: $\|\Theta\|_{\mathrm{HS}}^2 := \sum_{i,j=1}^{\infty} \langle f_i, \Theta g_j \rangle_k^2$, where the f_i's form a complete orthonormal system (CONS) for $\mathcal{H}_k$ and the g_j's form a CONS for $\mathcal{H}_c$.

4.1 AN UPPER CONFIDENCE-BOUND BASED ALGORITHM

The central idea is to maintain for each control $\mathbf{u}$, a confidence interval around the estimate $\langle \hat{\mu}_t, \hat{\vartheta}_t(\mathbf{u}) \rangle_k$ of the expected mean reward $\mathbb{E}_{P_\mathbf{u}}[f]$. Appropriate widths for the confidence intervals can be described in terms of the Mahalanobis norm of the control features $\phi_c(\mathbf{u})$ with respect to the regularized sample covariance matrix:

$$\sigma_t(\mathbf{u}) := \|\phi_c(\mathbf{u})\|_{(\mathbf{\Phi}_c(\mathbf{U}_t)\mathbf{\Phi}_c(\mathbf{U}_t)^\mathsf{T}+\eta\mathbf{I})^{-1}} \,.$$

An application of the Sherman-Morrison formula yields:

$$\sigma_t(\mathbf{u}) = \eta^{-1/2}\sqrt{c(\mathbf{u},\mathbf{u}) - \mathbf{c}_t(\mathbf{u})^\mathsf{T}(\mathbf{C}_t + \eta\mathbf{I})^{-1}\mathbf{c}_t(\mathbf{u})} \,. \tag{11}$$

Now, given a set of past observations $\mathcal{D}_{t-1} = \{(\mathbf{u}_i, \mathbf{x}_i, y_i)\}_{i=1}^{t-1}$, the following defines an upper confidence bound (UCB) acquisition function:

$$h(\mathbf{u}|\mathcal{D}_{t-1}) = \langle \hat{\mu}_{t-1}, \hat{\vartheta}_{t-1}(\mathbf{u}) \rangle_k + \beta_{t-1}\sigma_{t-1}(\mathbf{u}) \,, \tag{12}$$

where β_{t-1} is a parameter of the algorithm. The theoretical results in Section 5.2 will show that β_{t-1} can be set accordingly for $h(\mathbf{u}|\mathcal{D}_{t-1})$ to maintain a high-probability upper bound on $\mathbb{E}[f(\mathbf{x})|\mathbf{u}]$ for all $\mathbf{u} \in \mathcal{U}$.

Algorithm 1 presents the Conditional Mean Embeddings Upper Confidence Bound (CME-UCB) algorithm, which estimates the conditional mean embeddings as well as builds an UCB acquisition function $h(\mathbf{u}|\mathcal{D}_{t-1})$ using the estimates (Equation 12). At each iteration t, the algorithm selects a control $\mathbf{u}_t$ that maximises $h(\mathbf{u}|\mathcal{D}_{t-1})$ (line 2). Such a rule inherently trades off between exploration (picking points with high uncertainty) and exploitation (picking points with high expected reward) with the parameter β_{t-1} controlling this trade-off. In line 3, the objective function f is queried at some location $\mathbf{x}_t|\mathbf{u}_t \sim P_{\mathbf{u}_t}$. After the query is done, the algorithm is provided with an observation $y_t = f(\mathbf{x}_t) + \zeta_t$. In line 5, first the estimate $\hat{\mu}_t$ of f is updated with the new observation pair $(\mathbf{x}_t, y_t)$, and then $\boldsymbol{\mu}_t$, the vector evaluations of $\hat{\mu}_t$ at the observed states is computed. Finally in line 6, we update the estimate $\hat{\vartheta}_t(\mathbf{u})$ of the conditional mean embedding and confidence width $\sigma_t(\mathbf{u})$ with the augmented data point $(\mathbf{u}_t, \boldsymbol{\mu}_t)$. This process then repeats for a given number of iterations n.

Computational complexity: Computing $\langle \hat{\mu}_t, \hat{\vartheta}_t(\mathbf{u}) \rangle_k$ involves inversion and multiplication of t-by-t matrices (Equation 10), which take $O(t^3)$ time. Similarly, computing $\sigma_t(\mathbf{u})$ takes $O(t^3)$ time (Equation 11). Hence, the per-update time complexity of CME-UCB (Algorithm 1) is $O(t^3)$ which is no worse than standard BO algorithms, e.g., GP-UCB (Srinivas et al., 2010). However, using standard kernel approximation techniques like the *Nyström* approximation (Drineas and Mahoney, 2005) or the *random Fourier features* approximation (Rahimi and Recht, 2008) and efficient incremental update schemes (Gijsberts and Metta, 2013), the run-time complexity of our algorithm can be reduced to a $O(m^2)$ per-update cost, where $m \ll n$ is the dimension of feature approximations (Gijsberts and Metta, 2013).

Algorithm 1: Conditional Mean Embeddings Upper Confidence Bound (CME-UCB)

Input: $\mathcal{U}$: control space
n: total number of iterations

1 **for** $t \in \{1, \ldots, n\}$ **do**
2 $\quad \mathbf{u}_t = \operatorname{argmax}_{\mathbf{u}\in\mathcal{U}} \langle \hat{\mu}_{t-1}, \hat{\vartheta}_{t-1}(\mathbf{u}) \rangle_k + \beta_{t-1}\sigma_{t-1}(\mathbf{u})$
3 $\quad (\mathbf{x}_t, y_t) \leftarrow$ Sample f at $\mathbf{x}_t|\mathbf{u}_t \sim P_{\mathbf{u}_t}$
4 $\quad \mathcal{D}_t = \mathcal{D}_{t-1} \cup \{(\mathbf{u}_t, \mathbf{x}_t, y_t)\}$
5 $\quad$ Update $\hat{\mu}_t$ with $(\mathbf{x}_t, y_t)$ and compute $\boldsymbol{\mu}_t$
6 $\quad$ Update $\hat{\vartheta}_t(\mathbf{u})$ and $\sigma_t(\mathbf{u})$ with $(\mathbf{u}_t, \boldsymbol{\mu}_t)$

5 THEORETICAL RESULTS

This section presents our main theoretical results. We split them in two parts, first presenting novel approximation bounds for learning with conditional mean embeddings. We then derive the resulting regret bounds for the proposed CME-UCB algorithm.

5.1 ACTIVE LEARNING OF CONDITIONAL MEAN EMBEDDING

First we consider building a confidence ball in the RKHS $\mathcal{H}_k$ around the estimated conditional mean embedding $\hat{\vartheta}_t(\mathbf{u})$ such that the true embedding $\vartheta(\mathbf{u})$ lies in it with high probability. Equivalently, we focus on upper bounding the distance $\|\vartheta(\mathbf{u}) - \hat{\vartheta}_t(\mathbf{u})\|_k$ between the conditional mean embedding and its estimate as a function of the uncertainty $\sigma_t(\mathbf{u})$ around the estimate. Existing results (Song et al., 2010b) assume the data $\{\mathbf{u}_i, \mathbf{x}_i\}_{i=1}^t$ to be i.i.d. and, hence, not valid for our purpose.

To this end, we define a measure of the *information gain* about the conditional mean embeddings $\vartheta(\mathbf{u})$ as a function of the kernel c on $\mathcal{U}$ and the number of samples t:

$$\gamma_{c,t} = \sup_{\mathbf{U}\subset\mathcal{U}:|\mathbf{U}|=t} \frac{1}{2}\log\det\left(\mathbf{I} + \eta^{-1}\mathbf{C}_\mathbf{U}\right) \,, \tag{13}$$

where $\mathbf{C}_\mathbf{U} = [c(\mathbf{u},\mathbf{u}')]_{\mathbf{u},\mathbf{u}'\in\mathbf{U}}$ denotes the gram matrix computed at $\mathbf{U}$. Now, we derive a novel upper bound on $\|\vartheta(\mathbf{u}) - \hat{\vartheta}_t(\mathbf{u})\|_k$ under non i.i.d. data as a function of the uncertainty $\sigma_t(\mathbf{u})$.

Theorem 1. *Let $k(\mathbf{x},\mathbf{x}) \leq 1$ for all $\mathbf{x} \in \mathcal{X}$, $c(\mathbf{u},\mathbf{u}) \leq 1$ for all $\mathbf{u} \in \mathcal{U}$ and $\|\Theta\|_{\text{op}} \leq B$. Then, for any $\delta \in (0,1]$, with probability at least $1-\delta$, the following holds uniformly over all $t \geq 0$ and $\mathbf{u} \in \mathcal{U}$:*

$$\|\vartheta(\mathbf{u}) - \hat{\vartheta}_t(\mathbf{u})\|_k \leq \beta_{c,t}(\delta)\sigma_t(\mathbf{u}) \ ,$$

where $\beta_{c,t}(\delta) = B\sqrt{\eta} + 2\sqrt{2\left(\gamma_{c,t} + \log(1/\delta)\right)}$.

Proof. Since $\Theta : \mathcal{H}_c \to \mathcal{H}_k$ is a bounded linear operator, we have $g := \Theta^{\mathsf{T}} f \in \mathcal{H}_c$ for any $f \in \mathcal{H}_k$ (for the sake of completeness we provide a short proof for this result in the appendix, Lemma A.2). Hence,

$$\langle f, \vartheta(\mathbf{u})\rangle_k = \langle f, \Theta\phi_c(\mathbf{u})\rangle_k = \langle g, \phi_c(\mathbf{u})\rangle_c = g(\mathbf{u}) \ .$$

From Equation 9, we have:

$$\langle f, \hat{\vartheta}_t(\mathbf{u})\rangle_k = \mathbf{c}_t(\mathbf{u})^{\mathsf{T}}(\mathbf{C}_t + \eta\mathbf{I})^{-1}\mathbf{f}_t \ ,$$

where $\mathbf{f}_t = [f(\mathbf{x}_1), \ldots, f(\mathbf{x}_t)]^{\mathsf{T}}$. Let $\epsilon_t := k(\cdot, \mathbf{x}_t) - \vartheta(\mathbf{u}_t) = k(\cdot,\mathbf{x}_t) - \mathbb{E}[k(\cdot,\mathbf{x}_t)|\mathbf{u}_t]$. Then,

$$f(\mathbf{x}_t) = \langle f, \vartheta(\mathbf{u}_t)\rangle_k + \langle f, \epsilon_t\rangle_k = g(\mathbf{u}_t) + \langle f, \epsilon_t\rangle_k \ .$$

Note that $\|\mathbb{E}[k(\cdot,\mathbf{x}_t)|\mathbf{u}_t]\|_k \leq \mathbb{E}[k(\mathbf{x}_t,\mathbf{x}_t)|\mathbf{u}_t] \leq 1$, and hence $\|\epsilon_t\|_k \leq 2$. Let $\mathcal{F}_{t-1}$ be the sigma-algebra generated by the random variables $\{(\mathbf{u}_i,\mathbf{x}_i)\}_{i=1}^{t-1}$ and $\mathbf{u}_t$. Then $\langle f, \epsilon_t\rangle_k$ is zero-mean $2\|f\|_k$-sub-Gaussian given $\mathcal{F}_{t-1}$. For any $\delta \in (0,1]$, let $\alpha_{c,t}(\delta) = \|\Theta^{\mathsf{T}} f\|_c\sqrt{\eta} + 2\|f\|_k\sqrt{2\left(\gamma_{c,t} + \log(1/\delta)\right)}$. Then by Durand et al. (2018, Theorem 1), with probability at least $1-\delta$, uniformly over all $t \geq 0$ and $\mathbf{u} \in \mathcal{U}$, the following holds:

$$\left|\langle f, \vartheta(\mathbf{u})\rangle_k - \langle f, \hat{\vartheta}_t(\mathbf{u})\rangle_k\right| \leq \alpha_{c,t}(\delta)\sigma_t(\mathbf{u}) \ .$$

Now the result follows from the definition of operator norm and the fact that $\|\vartheta(\mathbf{u}) - \hat{\vartheta}_t(\mathbf{u})\|_k = \sup_{f\in\mathcal{H}_k:\|f\|_k\leq 1}\left|\langle f, \vartheta(\mathbf{u})\rangle_k - \langle f, \hat{\vartheta}_t(\mathbf{u})\rangle_k\right|$. □

Interpretation of the bound: In order to understand the growth of $\|\vartheta(\mathbf{u}) - \hat{\vartheta}_t(\mathbf{u})\|_k$ with the number of samples t, first we need to understand the behaviour of $\gamma_{c,t}$. Let $g : \mathcal{U} \to \mathbb{R}$ be a (random) function sampled from a zero-mean Gaussian process $\text{GP}(0,c)$ with covariance function c. Then $\gamma_{c,t}$ denotes the *maximum information gain* about g after t noisy observations obtained by passing g through an i.i.d. Gaussian channel $N(0,\eta)$, and it measures the reduction in the uncertainty of g after t noisy observations. $\gamma_{c,t}$ is a function of the kernel c and domain $\mathcal{U}$. If c is the squared exponential (SE) kernel and $\mathcal{U} \subset \mathbb{R}^d$ is compact and convex, then $\gamma_{c,t} = O\left((\log t)^{d+1}\right)$ (Srinivas et al., 2010). In this case, Theorem 1 implies that $\|\vartheta(\mathbf{u}) - \hat{\vartheta}_t(\mathbf{u})\|_k = O_p\left(\text{polylog}\, t\right)\sigma_t(\mathbf{u})$.[4] For the Matérn kernel $\gamma_{c,t} = O\left(t^{\frac{d(d+1)}{2\nu+d(d+1)}}\log t\right)$, and therefore $\|\vartheta(\mathbf{u}) - \hat{\vartheta}_t(\mathbf{u})\|_k = O_p\left(t^{\alpha}(\log t)^{1/2}\right)\sigma_t(\mathbf{u})$, where $\alpha < 1/2$. In fact, $\alpha < 1/4$ as long as $\nu > d(d+1)/2$.

Comparison with results under i.i.d. samples: Let $C_{\mathbf{uu}} = \mathbb{E}[\phi_c(\mathbf{u}) \otimes \phi_c(\mathbf{u})]$ and $C_{\mathbf{xu}} = \mathbb{E}[\phi_k(\mathbf{x}) \otimes \phi_c(\mathbf{u})]$ denote the (uncentered) covariance operator on $\mathcal{U}$ and cross-covariance operator from $\mathcal{U}$ to $\mathcal{X}$, respectively.[5] Then, under i.i.d. samples $\{\mathbf{u}_i,\mathbf{x}_i\}_{i=1}^t$, the distance (in RKHS norm) between the conditional mean embedding $C_{\mathbf{xu}}C_{\mathbf{uu}}^{-1}\phi_c(\mathbf{u})$ and its empirical estimate $\mathbf{\Phi}_k(\mathbf{X}_t)(\mathbf{C}_t + \eta t\mathbf{I})^{-1}\mathbf{c}_t(\mathbf{u})$ is of the order $O_p(1/\sqrt{t})$ (Song et al., 2010b, Theorem 1). Note that this estimator is different from $\hat{\vartheta}_t(\mathbf{u})$ in the sense that the regulariser used in the former expression is η multiplied by the number of samples t, whereas we use η as the regulariser (Equation 9). In that case, as evident from Equation 11, we can crudely upper bound $\sigma_t(\mathbf{u})$ by $1/\sqrt{\eta t}$ to recover the $O_p(1/\sqrt{t})$ scaling[6] of Song et al. (2010b) up to a polylog factor for the SE kernel and up to a t^{α} factor, $\alpha < 1/2$, for the Matérn kernel. We refrain from comparing with the results of Grünewälder et al. (2012a) since the latter assumes $\mathcal{H}_k$ to be a finite dimensional RKHS, which is generally not the case.

Application of the bound: For the purpose of this work, Theorem 1 will be used to build a confidence ellipsoid around the sample estimate of the conditional expectation $\mathbb{E}_{P_{\mathbf{u}}}[f]$. However, the applicability of Theorem 1 is much more general and might be of independent interest. We point out one such simple application. Let $\hat{P}_{\mathbf{u}}^t$ be a t sample empirical approximation to the conditional distribution $P_{\mathbf{u}}$ in the sense that $\mathbb{E}_{\hat{P}_{\mathbf{u}}^t}[f] = \langle f, \hat{\vartheta}_t(\mathbf{u})\rangle_k$ for any $f \in \mathcal{H}_k$. It then follows that:

$$\|\vartheta(\mathbf{u}) - \hat{\vartheta}_t(\mathbf{u})\|_k = \sup_{f\in\mathcal{H}_k:\|f\|_k\leq 1}\left|\mathbb{E}_{P_{\mathbf{u}}}[f] - \mathbb{E}_{\hat{P}_{\mathbf{u}}^t}[f]\right| \ ,$$

which corresponds to the *maximum mean discrepancy* (MMD) between distributions $P_{\mathbf{u}}$ and $\hat{P}_{\mathbf{u}}^t$. Thus, Theorem 1 can be used to provide a upper bound on the MMD between a (conditional) distribution and its estimate under non i.i.d. samples. The MMD, and more generally, kernel mean embeddings have been used in many applications particularly in kernel density estimation (Smola et al., 2007), two and one sample tests (Gretton et al.,

[4] $O_p(\cdot)$ hides constants and dependencies on $\log(1/\delta)$.

[5] $f \otimes g$ denotes the tensor product between $f \in \mathcal{H}_k$ and $g \in \mathcal{H}_c$, and for any $h \in \mathcal{H}_c$ satisfies $(f \otimes g)h = \langle g, h\rangle_c f$.

[6] Thanks to the smoothness of $\vartheta(\mathbf{u})$ (Equation 7), $\sigma_t(\mathbf{u})$ would decay faster than $O(1/\sqrt{t})$, yielding a better rate of convergence.

2012) and distributionally robust optimisation (Staib and Jegelka, 2019).

5.2 CONCENTRATION BOUND

Equipped with the confidence ball around $\hat{\vartheta}_t(\mathbf{u})$ constructed in Theorem 1, we now focus on building a confidence interval around the sample estimate $\langle \hat{\mu}_t, \hat{\vartheta}_t(\mathbf{u}) \rangle_k$ of $\langle f, \vartheta(\mathbf{u}) \rangle_k$, the expectation of f under the conditional distribution $P_{\mathbf{u}}$. However, first we need a confidence ball around the sample estimate $\hat{\mu}_t$ of the objective function f such that f lies in it with high probability. To this end, we describe the uncertainty around $\hat{\mu}_t(\mathbf{x})$ in terms of the Mahalanobis norm of the state features:

$$s_t(\mathbf{x}) = \lambda^{-1/2} \sqrt{k(\mathbf{x},\mathbf{x}) - \mathbf{k}_t(\mathbf{x})^\mathsf{T} (\mathbf{K}_t + \lambda \mathbf{I})^{-1} \mathbf{k}_t(\mathbf{x})} \, .$$

We also define a measure of the *information gain* about f as a function of the kernel k and number of samples t:

$$\gamma_{k,t} = \sup_{\mathbf{X} \subset \mathcal{X} : |\mathbf{X}| = t} \frac{1}{2} \log\det\left(\mathbf{I} + \lambda^{-1} \mathbf{K}_\mathbf{X}\right) \, , \qquad (14)$$

where $\mathbf{K}_\mathbf{X} = [k(\mathbf{x},\mathbf{x}')]_{\mathbf{x},\mathbf{x}' \in \mathbf{X}}$ denotes the gram matrix computed at $\mathbf{X}$. Lemma 2 provides a high probability confidence interval around $\hat{\mu}_t(\mathbf{x})$ with its width being proportional to $s_t(\mathbf{x})$, and is well known in the BO literature (Durand et al., 2018).

Lemma 2 (Durand et al. (2018))**.** *Let $k(\mathbf{x},\mathbf{x}) \leq 1$ for all $\mathbf{x} \in \mathcal{X}$, $\|f\|_k \leq b$ and ζ_t be σ_ζ-sub-Gaussian (Equation 2). Then, for any $\delta \in (0,1]$, with probability at least $1-\delta$, uniformly over all $t \geq 0$ and $\mathbf{x} \in \mathcal{X}$,*

$$|f(\mathbf{x}) - \hat{\mu}_t(\mathbf{x})| \leq \beta_{k,t}(\delta) s_t(\mathbf{x}) \, ,$$

where $\beta_{k,t}(\delta) = b\sqrt{\lambda} + \sigma_\zeta \sqrt{2\left(\gamma_{k,t} + \log(1/\delta)\right)}$.

Now, in Lemma 3, we build a high probability confidence interval around $\langle \hat{\mu}_t, \hat{\vartheta}_t(\mathbf{u}) \rangle_k$ with its width proportional to $\sigma_t(\mathbf{u})$, the uncertainty around the approximate embedding $\hat{\vartheta}_t(\mathbf{u})$. However, the width is inflated accordingly to account for the uncertainty around $\hat{\mu}_t$.

Lemma 3. *Let $k(\mathbf{x},\mathbf{x}) \leq 1$ for all $\mathbf{x} \in \mathcal{X}$, $c(\mathbf{u},\mathbf{u}) \leq 1$ for all $\mathbf{u} \in \mathcal{U}$, $\|f\|_k \leq b$, $\|\Theta\|_{\mathrm{op}} \leq B$ and ζ_t be σ_ζ-sub-Gaussian (Equation 2). For any $\delta \in (0,1]$, let $\beta_{c,t}(\delta)$ and $\beta_{k,t}(\delta)$ be as given in Theorem 1 and Lemma 2, respectively. Then, with probability at least $1-\delta$, the following holds uniformly over all $t \geq 0$ and $\mathbf{u} \in \mathcal{U}$:*

$$\left| \langle f, \vartheta(\mathbf{u}) \rangle_k - \langle \hat{\mu}_t, \hat{\vartheta}_t(\mathbf{u}) \rangle_k \right| \leq \beta_t \sigma_t(\mathbf{u}) \, ,$$

where $\beta_t := \beta_t(\delta) = b\beta_{c,t}(\delta/2) + \beta_{k,t}(\delta/2)\sqrt{2\gamma_{k,t}}$.

Proof. By an application of triangle inequality, we have $\left| \langle f, \vartheta(\mathbf{u}) \rangle_k - \langle \hat{\mu}_t, \hat{\vartheta}_t(\mathbf{u}) \rangle_k \right| \leq \left| \langle f, \vartheta(\mathbf{u}) - \hat{\vartheta}_t(\mathbf{u}) \rangle_k \right| + \left| \langle f - \hat{\mu}_t, \hat{\vartheta}_t(\mathbf{u}) \rangle_k \right|$. For the first term, by the Cauchy-Schwartz inequality and Theorem 1, with probability at least $1-\delta$, the following holds for all $t > 0$ and $\mathbf{u} \in \mathcal{U}$:

$$\left| \langle f, \vartheta(\mathbf{u}) - \hat{\vartheta}_t(\mathbf{u}) \rangle_k \right| \leq b\beta_{c,t}(\delta)\sigma_t(\mathbf{u}) \, , \qquad (15)$$

since $\|f\|_k \leq b$. For the second term, let $\mathbf{f}_t = [f(\mathbf{x}_1), \ldots, f(\mathbf{x}_t)]^\mathsf{T}$. From Equation 9, we then have

$$\begin{aligned} \left| \langle f - \hat{\mu}_t, \hat{\vartheta}_t(\mathbf{u}) \rangle_k \right| &= \left| \mathbf{c}_t(\mathbf{u})^\mathsf{T} (\mathbf{C}_t + \eta \mathbf{I})^{-1} (\mathbf{f}_t - \boldsymbol{\mu}_t) \right| \\ &\leq \|(\mathbf{C}_t + \eta \mathbf{I})^{-1} \mathbf{c}_t(\mathbf{u})\|_2 \|\mathbf{f}_t - \boldsymbol{\mu}_t\|_2 \\ &= \|\boldsymbol{\Phi}_c(\mathbf{U}_t)^\mathsf{T} \mathbf{V}_t^{-1} \phi_c(\mathbf{u})\|_2 \|\mathbf{f}_t - \boldsymbol{\mu}_t\|_2, \end{aligned}$$

where $\mathbf{V}_t = (\boldsymbol{\Phi}_c(\mathbf{U}_t)\boldsymbol{\Phi}_c(\mathbf{U}_t)^\mathsf{T} + \eta \mathbf{I})^{-1}$. It is easy to see that $\|\boldsymbol{\Phi}_c(\mathbf{U}_t)^\mathsf{T} \mathbf{V}_t^{-1} \phi_c(\mathbf{u})\|_2 \leq \sigma_t(\mathbf{u})$, since $\|\boldsymbol{\Phi}_c(\mathbf{U}_t)^\mathsf{T} \mathbf{V}_t^{-1/2}\|_{\mathrm{op}} \leq 1$. Now by Lemma 2 and monotonicity of $\beta_{k,t}$, we have with probability at least $1-\delta$:

$$\|\mathbf{f}_t - \boldsymbol{\mu}_t\|_2 \leq \beta_{k,t}(\delta) \sqrt{\sum_{i=1}^t s_i^2(\mathbf{x}_i)} \leq \beta_{k,t}(\delta)\sqrt{2\gamma_{k,t}} \, ,$$

where the final step follows from $\sum_{i=1}^t s_i^2(\mathbf{x}_i) \leq \log\det\left(\lambda^{-1}\mathbf{K}_t + \mathbf{I}\right)$ (Lemma A.1 in the Appendix) and the definition of $\gamma_{k,t}$ (Equation 14). Therefore, with probability at least $1-\delta$, for all $t > 0$ and $\mathbf{u} \in \mathcal{U}$:

$$\left| \langle f - \hat{\mu}_t, \hat{\vartheta}_t(\mathbf{u}) \rangle_k \right| \leq \beta_{k,t}(\delta)\sqrt{2\gamma_{k,t}}\sigma_t(\mathbf{u}) \, . \qquad (16)$$

Now the result follows by combining Equation 15 and Equation 16, and taking an union bound. □

5.3 REGRET BOUND OF CME-UCB

In this section, we derive an upper bound on the cumulative regret of CME-UCB (Algorithm 1).

Theorem 4. *Fix any $\delta \in (0,1]$. Then, under the same hypothesis of Lemma 3, CME-UCB with β_t set as in Lemma 3, enjoys, with probability at least $1-\delta$, the regret bound:*

$$R_n \leq 2\beta_n \sqrt{2(1+1/\eta)\gamma_{c,n}\, n} \, .$$

Proof. Let $\left| \langle f, \vartheta(\mathbf{u}) \rangle_k - \langle \hat{\mu}_t, \hat{\vartheta}_t(\mathbf{u}) \rangle_k \right| \leq \beta_t \sigma_t(\mathbf{u})$ for all $t \geq 0$ and $\mathbf{u} \in \mathcal{U}$. Then the instantaneous regret at time $t \geq 1$ is:

$$\begin{aligned} r_t &:= \mathbb{E}[f(\mathbf{x}) | \mathbf{u}^\star] - \mathbb{E}[f(\mathbf{x}) | \mathbf{u}_t] \\ &= \langle f, \vartheta(\mathbf{u}^\star) \rangle_k - \langle f, \vartheta(\mathbf{u}_t) \rangle_k \\ &\leq \langle \hat{\mu}_{t-1}, \hat{\vartheta}_{t-1}(\mathbf{u}^\star) \rangle_k + \beta_{t-1}\sigma_{t-1}(\mathbf{u}^\star) - \langle f, \vartheta(\mathbf{u}_t) \rangle_k \\ &\leq \langle \hat{\mu}_{t-1}, \hat{\vartheta}_{t-1}(\mathbf{u}_t) \rangle_k + \beta_{t-1}\sigma_{t-1}(\mathbf{u}_t) - \langle f, \vartheta(\mathbf{u}_t) \rangle_k \\ &\leq 2\beta_{t-1}\sigma_{t-1}(\mathbf{u}_t) \, , \end{aligned}$$

where the second inequality is due the choice of $\mathbf{u}_t$ (line 2 of Algorithm 1). Now, by Lemma 3, with probability at least $1-\delta$, we have:

$$R_n \leq \sum_{t=1}^{n} 2\beta_{t-1}\sigma_{t-1}(\mathbf{u}_t) \leq 2\beta_n \sqrt{n \sum_{t=1}^{n} \sigma_{t-1}^2(\mathbf{u}_t)}\,,$$

where the last step is due to the monotonicity of β_t and the Cauchy-Schwartz inequality. Now the result follows from the definition of $\gamma_{c,t}$ (Equation 13) along with the identities $\sigma_{t-1}^2(\mathbf{u}) \leq (1+1/\eta)\sigma_t^2(\mathbf{u})$ and $\sum_{t=1}^{n} \sigma_t^2(\mathbf{u}_t) = \log\det\left(\lambda^{-1}\mathbf{C}_n + \mathbf{I}\right)$ (Refer to Lemma A.1 in the Appendix.). □

Theorem 4 implies that the cumulative regret of CME-UCB is $O_p\left(\gamma_{k,n}\sqrt{\gamma_{c,n}\, n} + \gamma_{c,n}\sqrt{n}\right)$. If k and c are both squared exponential kernels, the regret is of the order $\tilde{O}(\sqrt{n})$, and hence sublinear in n.[7]

Comparison with the GP-UCB algorithm: For $g := \Theta^\mathsf{T} f$, we can write the observation y_t as a noisy sample of $g(\mathbf{u}_t)$ corrupted by zero-mean noise $\langle f, \epsilon_t\rangle_k + \zeta_t$, which is $\sqrt{\sigma_\zeta^2 + 4b^2}$ sub-Gaussian. Then, instead of estimating the conditional mean embedding, one can directly run GP-UCB (Chowdhury and Gopalan, 2017) as

$$\mathbf{u}_t = \operatorname{argmax}_{\mathbf{u}\in\mathcal{U}} \tilde{\mu}_{t-1}(\mathbf{u}) + \tilde{\beta}_{t-1}\sigma_{t-1}(\mathbf{u})\,,$$

with $\tilde{\mu}_t(\mathbf{u}) = \mathbf{c}_t(\mathbf{u})^\mathsf{T}(\mathbf{C}_t + \eta\mathbf{I})^{-1}\mathbf{y}_t$ and $\tilde{\beta}_t = bB\sqrt{\eta} + \sqrt{\sigma_\zeta^2 + 4b^2}\sqrt{2(\gamma_{c,t} + \log(1/\delta))}$, and achieve a regret bound of $O_p\left(\tilde{\beta}_n\sqrt{\gamma_{c,n}\, n}\right)$. Since $\beta_t = O_p\left(\sigma_\zeta\gamma_{k,t} + b(\sqrt{\gamma_{k,t}} + \sqrt{\gamma_{c,t}})\right)$ and $\tilde{\beta}_t = O_p\left(\sqrt{(\sigma_\zeta^2 + b^2)\gamma_{c,t}}\right)$, Theorem 4 is tighter than that of GP-UCB as long as $b \ll \sigma_\zeta$ and $\gamma_{k,t} \ll \sqrt{\gamma_{c,t}}$. Particularly, if the dimension D of the state space is much smaller than the control space dimension d (e.g. $D \ll d/2$ for the SE kernel), then there is a clear advantage of estimating the conditional mean embedding using the observed states as an intermediate step, as proposed in Algorithm 1. The learnt embedding also encodes knowledge of the dynamics of the stochastic query process, which can be applied to future tasks.

6 APPLICATION TO LIKELIHOOD-FREE INFERENCE

In this section, we discuss an application of Theorem 1, and the key insights it brings, to the likelihood-free inference problem. The goal is to estimate parameters $\mathbf{u}$ of a physical system according to observed data, represented by summary statistics $\mathbf{x}_o$. The forward model of the system is approximated by a simulation model $\mathbf{x}|\mathbf{u} \sim P_\mathbf{u}$ based on the given parameter settings $\mathbf{u}$. In this case, the likelihood $p(\mathbf{x}_o|\mathbf{u})$ is given by:

$$p(\mathbf{x}_o|\mathbf{u}) = \int_{\mathbf{x}\in\mathcal{X}} p(\mathbf{x}_o|\mathbf{x})p(\mathbf{x}|\mathbf{u})\,\mathrm{d}\mathbf{x} \tag{17}$$

The integral above is often intractable due to the marginalisation over simulator outputs. Simulations can also be expensive and time consuming, difficulting inference even further. As proposed by Hsu and Ramos (2019), for a symmetric simulator-output likelihood, i.e. $p(\mathbf{x}_o|\mathbf{x}) = p(\mathbf{x}|\mathbf{x}_o)$, such as the densities in the exponential family, we can set $p(\mathbf{x}_o|\mathbf{x}) =: k(\mathbf{x}_o, \mathbf{x})$, for any $\mathbf{x}, \mathbf{x}_o \in \mathcal{X}$. In this case, we have:

$$p(\mathbf{x}_o|\mathbf{u}) = \mathbb{E}_\mathbf{x}[k(\mathbf{x}_o, \mathbf{x})|\mathbf{u}] = \langle \phi_k(\mathbf{x}_o), \vartheta(\mathbf{u})\rangle_k \tag{18}$$

The method proposed by Hsu and Ramos (2019) still used i.i.d. data $\{\mathbf{u}_t, \mathbf{x}_t\}_{t=1}^n$. With our results, however, we can formulate an informative sampling approach.

Let $\psi_{\hat{P}_\mathbf{u}^t}(\mathbf{x}_o) := \langle \phi_k(\mathbf{x}_o), \hat{\vartheta}_t(\mathbf{u})\rangle_k = \mathbf{k}_t(\mathbf{x}_o)^\mathsf{T}(\mathbf{C}_t + \eta\mathbf{I})^{-1}\mathbf{c}_t(\mathbf{u})$. Then, applying Theorem 1 to Equation 18, with probability greater than $1-\delta$, the following concentration bound holds:

$$\forall \mathbf{u} \in \mathcal{U}\,, \forall t \geq 0,\ |p(\mathbf{x}_o|\mathbf{u}) - \psi_{\hat{P}_\mathbf{u}^t}(\mathbf{x}_o)| \leq \beta_{c,t}(\delta)\sigma_t(\mathbf{u})\,, \tag{19}$$

since $\|\phi_k(\mathbf{x}_o)\|_k = \sqrt{k(\mathbf{x}_o, \mathbf{x}_o)} \leq 1$, under our settings. At every iteration, an algorithm can select:

$$\mathbf{u}_t \in \operatorname*{argmax}_{\mathbf{u}\in\mathcal{U}} \psi_{\hat{P}_\mathbf{u}^{t-1}}(\mathbf{x}_o) + \beta_{c,t-1}\sigma_{t-1}(\mathbf{u}) \tag{20}$$

Following this strategy, the algorithm approaches the maximum likelihood estimator $\mathbf{u}^* \in \operatorname{argmax}_{\mathbf{u}\in\mathcal{U}} p(\mathbf{x}_o|\mathbf{u})$ with regret bounded by Theorem 4, setting $\beta_t := \beta_{c,t}, \forall t \geq 0$.

To demonstrate the approach, Figure 1 presents experimental results on learning the likelihod $p(\mathbf{x}|\mathbf{u})$ via

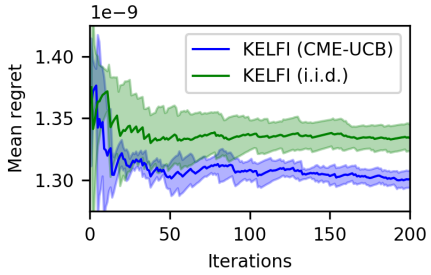

Figure 1: Regret comparison for KELFI with the UCB strategy vs. i.i.d. data. Results are averaged over 5 trials. The plot shows the mean regret R_n/n with respect to the maximum likelihood estimator for the simulator parameters. Shaded areas correspond to ± 1 standard deviation.

[7] $\tilde{O}(\cdot)$ hides constants and log factors.

KELFI (Hsu and Ramos, 2019) for the Lotka-Volterra simulator. For this experiment, we use a dataset with 10,000 i.i.d. pairs of simulator parameters $\mathbf{u}$ and observation summary statistics $\mathbf{x}$ to form the control and state spaces. We compare KELFI run with i.i.d. data, as originally proposed, against data selected by CME-UCB. The maximum likelihood estimator is computed based on a CME model fit with the entire dataset. A bound for $\|\Theta\|_{\text{op}}$ is also computed based on the same full-data CME model (see Lemma B.1). As Figure 1 shows, KELFI run with CME-UCB is able to quickly learn a likelihood model while approaching the maximum likelihood estimator. CME-UCB is able to find the most informative data points within the first 50 iterations.

7 A PRACTICAL IMPROVEMENT

This section presents a modification to the CME-UCB algorithm[8] which renders significant improvements over GP-UCB for optimisation problems involving unknown conditional distributions. We start by presenting theoretical results for the derivation of the improved CME-UCB algorithm and proceed with experimental results.[9]

7.1 A REFINED CME-UCB

In its proposed version, the upper bound in CME-UCB (Lemma 3) is still large when compared to the traditional GP-UCB. As a result, performance improvements are marginal in practical applications, as we demonstrate in the next section. However, we are able to further tighten the upper confidence bound in our main result, which leads to an improvement over GP-UCB. Proofs are deferred to Appendix C in the supplementary material.

Lemma 5. *For any $\delta \in (0,1]$, with probability at least $1-\delta$, uniformly over all $t \geq 0$ and $\mathbf{u} \in \mathcal{U}$,*

$$\left| \langle f, \hat{\vartheta}_t(\mathbf{u}) \rangle_k - \langle \hat{\mu}_t, \hat{\vartheta}_t(\mathbf{u}) \rangle_k \right| \leq \beta_{k,t}(\delta) s_t(\hat{P}^t_{\mathbf{u}}) \, .$$

The predictive variance $s_t^2(\hat{P}^t_{\mathbf{u}})$ on the approximate conditional $\hat{P}^t_{\mathbf{u}}$ is defined as:

$$s_t^2(\hat{P}^t_{\mathbf{u}}) := \lambda^{-1} \mathbf{v}_t(\mathbf{u})^{\mathsf{T}} k_t(\mathbf{X}_t, \mathbf{X}_t) \mathbf{v}_t(\mathbf{u}) \, , \tag{21}$$

where $\mathbf{v}_t(\mathbf{u}) := (\mathbf{C}_t + \eta \mathbf{I})^{-1} \mathbf{c}_t(\mathbf{u})$ and $k_t(\mathbf{X}_t, \mathbf{X}_t)$ is a GP predictive covariance matrix on the observed states:

$$\begin{aligned} k_t(\mathbf{X}_t, \mathbf{X}_t) &= \mathbf{K}_t - \mathbf{K}_t (\mathbf{K}_t + \lambda \mathbf{I})^{-1} \mathbf{K}_t \\ &= \lambda \mathbf{K}_t (\mathbf{K}_t + \lambda \mathbf{I})^{-1} \, . \end{aligned} \tag{22}$$

[8]The results presented in this section have been obtained post-submission.

[9]Code available at: `https://github.com/rafaol/active-learning-conditional-mean-embeddings`

Combining Lemma 5 with Theorem 1, we obtain the following refinement over Lemma 3.

Proposition 6. *For any $\delta \in (0,1]$, let $\beta_{c,t}(\delta)$ and $\beta_{k,t}(\delta)$ be as given in Theorem 1 and Lemma 2, respectively. Then, with probability at least $1-\delta$, the following holds:*

$$\forall t \geq 0, \ \forall \mathbf{u} \in \mathcal{U}, \ \left| \langle f, \vartheta(\mathbf{u}) \rangle_k - \langle \hat{\mu}_t, \hat{\vartheta}_t(\mathbf{u}) \rangle_k \right| \leq \beta_t(\mathbf{u}) \, ,$$

where $\beta_t(\mathbf{u}) := b\beta_{c,t}(\delta/2)\sigma_t(\mathbf{u}) + \beta_{k,t}(\delta/2) s_t(\hat{P}^t_{\mathbf{u}})$.

Using Proposition 6, we define a refined CME-UCB rule:

$$\mathbf{u}_t = \text{argmax}_{\mathbf{u} \in \mathcal{U}} \langle \hat{\mu}_{t-1}, \hat{\vartheta}_{t-1}(\mathbf{u}) \rangle_k + \beta_{t-1}(\mathbf{u}) \, .$$

We now derive a regret upper bound for this refined version of the CME-UCB algorithm, which we refer to as improved CME-UCB, or I-CME-UCB.

Theorem 7. *Fix any $\delta \in (0,1]$. Then, under the same hypothesis of Proposition 6, I-CME-UCB, enjoys, with probability at least $1-\delta$, the regret bound:*

$$R_n \leq 2 \left(b\beta_{c,n}(\delta/2) + \beta_{k,n}(\delta/2) \right) \sqrt{2(1+1/\eta)\gamma_{c,n} \, n} \, .$$

Comparison with GP-UCB. Theorem 7 implies an improved regret bound of order $O_p\left(\sigma_\zeta \sqrt{\gamma_{k,n}\gamma_{c,n} \, n} + b\gamma_{c,n}\sqrt{n}\right)$ for I-CME-UCB. Now GP-UCB, when applied to our setting, achieves a regret bound of order $O_p\left((\sigma_\zeta + b)\gamma_{c,n}\sqrt{n}\right)$. Therefore this improved bound for I-CME-UCB is tighter than GP-UCB as long as $\gamma_{k,n} \leq \gamma_{c,n}$.

7.2 EXPERIMENTAL RESULTS

We assess the performance of CME-UCB in comparison to GP-UCB on a synthetic toy example. For this experiment, we generate objective functions $f = \sum_{i=1}^m \alpha_i k(\cdot, \mathbf{x}_i) \in \mathcal{H}_k$ by sampling each $\mathbf{x}_i$ from a uniform distribution $U[0,1]$ and α_i from a Gaussian $\boldsymbol{\alpha} \sim N(\mathbf{0}, \mathbf{K})$, where $[\mathbf{K}]_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$. To ensure we find the global optimum of the acquisition function, the control space is finite $\mathcal{U} := \{\mathbf{u}_i\}_{i=1}^{n_{\mathbf{u}}}$. State conditional distributions were synthesised as Gaussians $P_{\mathbf{u}} := N(\hat{\mathbf{x}}(\mathbf{u}), \boldsymbol{\Sigma})$, where $\hat{\mathbf{x}}(\mathbf{u}) := \mathbf{A}\mathbf{u}$ with a randomly generated $\mathbf{A}$. The state kernel k is set as the squared exponential, while the control kernel is set as a Matérn with smoothness parameter $\nu = 2.5$ (Rasmussen and Williams, 2006). The algorithms are configured with $\delta = 0.2$ and the norm bounds on f and Θ are computed as the exact norms, which is possible due to a finite control space $\mathcal{U}$ (see Lemma B.1 in the supplement).

Figure 2 presents performance results for this experiment. As the plot shows, I-CME-UCB presents a significant improvement in terms of regret when compared

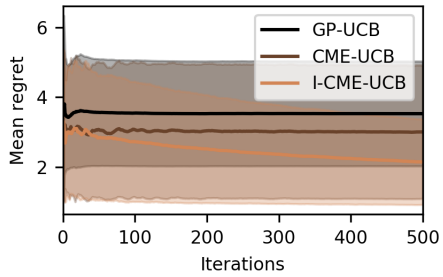

Figure 2: Regret comparison with CME-UCB, its improved version (I-CME-UCB) and GP-UCB on the toy experiment. Results are averaged over 10 trials. The shaded areas correspond to ± 1 standard deviation.

to CME-UCB. While CME-UCB still outperforms GP-UCB, its performance gains are only marginal when compared to I-CME-UCB. The main reason for this improvement is due to a less exploratory behaviour by I-CME-UCB coming from the tighter UCB.

8 CONCLUSION

In this paper we propose a novel method to optimise the conditional expectation of functions over a state space given a set of control variables, without any direct control over the states. Under the RKHS regularity assumption over the objective function, we first estimate the conditional mean embeddings and derive novel approximation bounds for these estimates under non i.i.d. samples. We then make use of these embeddings to design a Bayesian optimisation algorithm for maximising the conditional expectation and provide its regret bound. In terms of applications, we demonstrate that the proposed theoretical results lead to novel algorithms in likelihood-free inference. It may be possible to apply our method to other problems such as two sample test, kernel density estimation and distributionally robust optimisation, and we leave those as possible future work (see Appendix D for a discussion on applications to reinforcement learning). Another important future direction is adapting CME-UCB to efficiently choose a batch of controls (Desautels et al., 2014; Contal et al., 2013).

Acknowledgements

The authors are grateful to anonymous reviewers for providing useful comments.

References

Hans Georg Beyer and Bernhard Sendhoff. Robust optimization - A comprehensive survey. *Computer Methods in Applied Mechanics and Engineering*, 196(33-34):3190–3218, 2007.

Poompol Buathong, David Ginsbourger, and Tipaluck Krityakierne. Kernels over sets of finite sets using rkhs embeddings, with application to bayesian (combinatorial) optimization. *arXiv preprint arXiv:1910.04086*, 2019.

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.

Emile Contal, David Buffoni, Alexandre Robicquet, and Nicolas Vayatis. Parallel gaussian process optimization with upper confidence bound and pure exploration. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 225–240. Springer, 2013.

Marc Peter Deisenroth, Gerhard Neumann, and Jan Peters. A Survey on Policy Search for Robotics. *Foundations and Trends in Robotics*, 2(1-2):1–142, 2011.

Thomas Desautels, Andreas Krause, and Joel W Burdick. Parallelizing exploration-exploitation tradeoffs in gaussian process bandit optimization. *Journal of Machine Learning Research*, 15:3873–3923, 2014.

Petros Drineas and Michael W Mahoney. On the nyström method for approximating a gram matrix for improved kernel-based learning. *Journal of Machine Learning Research*, 6:2153–2175, 2005.

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.

Seth Flaxman, Dino Sejdinovic, John P. Cunningham, and Sarah Filippi. Bayesian Learning of Kernel Embeddings. In *Proceedings of the Thirty-Second Conference on Uncertainty in Artificial Intelligence*, UAI'16, pages 182–191, Arlington, Virginia, United States, 2016. AUAI Press.

Kenji Fukumizu, Arthur Gretton, Xiaohai Sun, and Bernhard Schölkopf. Kernel measures of conditional dependence. In *Advances in neural information processing systems*, pages 489–496, 2008.

Kenji Fukumizu, Francis R Bach, Michael I Jordan, et al. Kernel dimension reduction in regression. *The Annals of Statistics*, 37(4):1871–1905, 2009.

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

Arthur Gretton, Karsten M Borgwardt, Malte J Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *Journal of Machine Learning Research*, 13:723–773, 2012.

Steffen Grünewälder, Guy Lever, Luca Baldassarre, Sam Patterson, Arthur Gretton, and Massimilano Pontil. Conditional mean embeddings as regressors. In *Proceedings of the 29th International Coference on International Conference on Machine Learning*, pages 1803–1810, 2012a.

Steffen Grünewälder, Guy Lever, Luca Baldassarre, Massimilano Pontil, and Arthur Gretton. Modelling transition dynamics in mdps with rkhs embeddings. In *Proceedings of the 29th International Coference on International Conference on Machine Learning*, pages 1603–1610, 2012b.

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.

Kelvin Hsu and Fabio Ramos. Bayesian Learning of Conditional Kernel Mean Embeddings for Automatic Likelihood-Free Inference. In *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics (AISTATS)*, Naha, Okinawa, Japan, 2019.

Roman Marchant and Fabio Ramos. Bayesian Optimisation for Intelligent Environmental Monitoring. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, October 2012.

Krikamol Muandet, Kenji Fukumizu, Bharath Sriperumbudur, and Bernhard Schölkopf. Kernel Mean Embedding of Distributions: A Review and Beyond. *arXiv*, 2016.

Rafael Oliveira, Lionel Ott, and Fabio Ramos. Bayesian optimisation under uncertain inputs. In *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics (AISTATS)*, Naha, Okinawa, Japan, 2019.

George Papamakarios, David C. Sterratt, and Iain Murray. Sequential Neural Likelihood: Fast Likelihood-free Inference with Autoregressive Flows. In *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2019.

Ali Rahimi and Benjamin Recht. Random features for large-scale kernel machines. In *Advances in neural information processing systems*, pages 1177–1184, 2008.

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

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

Alex Smola, Arthur Gretton, Le Song, and Bernhard Schölkopf. A hilbert space embedding for distributions. In *International Conference on Algorithmic Learning Theory*, pages 13–31. Springer, 2007.

Jasper Snoek, Hugo Larochelle, and Ryan P. Adams. Practical bayesian optimization of machine learning algorithms. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 2951–2959. Curran Associates, Inc., 2012.

Le Song, Jonathan Huang, Alex Smola, and Kenji Fukumizu. Hilbert space embeddings of conditional distributions with applications to dynamical systems. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 961–968, 2009.

Le Song, Byron Boots, Sajid M Siddiqi, Geoffrey Gordon, and Alex Smola. Hilbert space embeddings of hidden markov models. In *Proceedings of the 27th International Conference on International Conference on Machine Learning*, pages 991–998, 2010a.

Le Song, Arthur Gretton, and Carlos Guestrin. Nonparametric tree graphical models. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pages 765–772, 2010b.

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.

Bharath K. Sriperumbudur, Kenji Fukumizu, and Gert R. G. Lanckriet. Universality, Characteristic Kernels and RKHS Embedding of Measures. *Journal of Machine Learning Research (JMLR)*, 12:2389–2410, 2011.

Matthew Staib and Stefanie Jegelka. Distributionally robust optimization and generalization in kernel methods. In *Advances in Neural Information Processing Systems*, pages 9131–9141, 2019.

Ingo Steinwart and Andreas Christmann. *Support Vector Machines*, chapter 4, pages 110–163. Springer New York, New York, NY, 2008.

Ngo Anh Vien and Marc Toussaint. Bayesian Functional Optimization. In *AAAI Conference on Artificial Intelligence*, pages 4171–4178, New Orleans, LA, USA, 2018.

Aaron Wilson, Alan Fern, and Prasad Tadepalli. Using trajectory data to improve bayesian optimization for reinforcement learning. *Journal of Machine Learning Research*, 15:253–282, 2014.