Distributionally-Aware Kernelized Bandit Problems for Risk Aversion

Sho Takemori¹

Abstract

The kernelized bandit problem is a theoretically justified framework and has solid applications to various fields. Recently, there is a growing interest in generalizing the problem to the optimization of risk-averse metrics such as Conditional Value-at-Risk (CVaR) or Mean-Variance (MV). However, due to the model assumption, most existing methods need explicit design of environment random variables and can incur large regret because of possible high dimensionality of them. To address the issues, in this paper, we model the environment using a family of the output distributions (or more precisely, probability kernel) and Kernel Mean Embeddings (KME), and provide novel UCB-type algorithms for CVaR and MV. Moreover, we provide algorithm-independent lower bounds for CVaR in the case of Matérn kernels, and propose a nearly optimal algorithm. Furthermore, we empirically verify our theoretical result in synthetic environments, and demonstrate that our proposed method significantly outperforms a baseline in many cases.

1. Introduction

The kernelized bandit problem, or Bayesian optimization problem (Srinivas et al., 2010; Chowdhury & Gopalan, 2017) is an online black-box optimization problem and has solid applications to various fields including material design, drug discovery, hyperparameter optimization, and recommender system (Frazier, 2018; Shahriari et al., 2015). Although, in the conventional kernelized bandit setting, the objective is to maximize the mean function, it may not be appropriate to optimize it in high-stake settings. For example, to optimize hyperparameters of a machine learning algorithm for medical domain, healthcare, or finance, one might optimize the mean performance with small variance (Mean-Variance optimization), or want to optimize more sophisticated risk averse metrics such as Conditional Valueat-Risk (CVaR). Recently, there is a growing interest in generalizing the optimization of the mean function to that of risk-averse metrics such as CVaR, Value-at-Risk (VaR), and Mean-Variance (MV) (Cakmak et al., 2020; Nguyen et al., 2021b;a; Iwazaki et al., 2021; Makarova et al., 2021). Here, for a distribution ρ , the Mean-Variance is defined as $\mathbf{E}[\rho] - c\mathbf{V}[\rho]$, where c > 0 is a parameter of the metric. CVaR is a metric for risk-aversion established in finance (Rockafellar et al., 2000) and given as the conditional expectation $\mathbf{E}_{y \sim \rho} [y \mid y \leq y_{\alpha}]$, where y_{α} is the α -quantile and $\alpha \in (0,1]$ is a parameter called the risk level. In this paper, we consider kernelized bandit problems for risk-averse metrics. We mainly focus on CVaR and MV, especially on CVaR, because CVaR has nice theoretical properties such as coherency or continuity on α (Acerbi & Tasche, 2002).

In most existing work on kernelized bandit problems for risk-averse metrics, the authors model the output y using a function y = f(x, W) with two kinds of variables, where x is the input variable that can be controlled by algorithms, and W is a random variable called the environment random variable that accounts for randomness of the output yand cannot be controlled by the algorithms (Marzat et al., 2013). In this paper, we call such a model the environmentvariable-function (EVF) model. However, the EVF model assumption can cause some restrictions and drawbacks. For instance, recently, Nguyen et al. (2021a) considered kernelized bandit problems for CVaR under the conditions that the distribution of W is known, the leaner can select W in the optimization procedure. If the environment is the real world or a complicated simulator, these conditions are too restrictive, since it is necessary to design the environment random variable that affects the output y, and it is assumed that the leaner can control/observe it during the optimization procedure. Moreover, since the regret upper bound in Nguyen et al. (2021a) is provided by using the maximum information gain (Srinivas et al., 2010) of a function with respect to (x, W), the regret upper bound increases if the dimension of W increases even if that of x is moderate.

Contributions

In this paper, by embedding the output distribution $\rho(x)$ at a point x via Kernel Mean Embedding (KME) (Muandet et al., 2017) and motivated by the regression interpretation of

¹Fujitsu Ltd., Kawasaki, Japan. Correspondence to: Sho Takemori <takemori.sho@fujitsu.com>.

Proceedings of the 39th International Conference on Machine Learning, Baltimore, Maryland, USA, PMLR 162, 2022. Copyright 2022 by the author(s).

Conditional Mean Embeddings (CME) (Grünewälder et al., 2012), we model a family of the output distributions (or more precisely, probability kernel) $\{\rho(x)\}_x$ by a bounded linear operator between two Reproducing Kernel Hilbert Spaces (RKHS). Although this model assumption is closely related to CME (Song et al., 2009; Park & Muandet, 2021), to distinguish the differences in theoretical conditions, we term our model the Probability Kernel Embedding (PKE) model. We provide more details in §3. Our contributions are summarized as follows:

(i) We provide a formulation of the PKE model (§3), and propose a novel UCB-type algorithm called CVPKE-UCB for the kernelized CVaR bandit problem without the aforementioned drawbacks of the method based on the EVF model, and provide a sublinear regret upper bound in the case of discrete output distributions (§4).

(ii) To show the PKE model is sufficiently generic to handle metrics other than CVaR, we also provide a kernelized bandit algorithm called MVPKE-UCB for MV (§5), which achieves a similar objective as recent work (Makarova et al., 2021).

(iii) Furthermore, we provide a lower bound of our cumulative regret for CVaR optimization in the case of Matérn kernels (§6).

(iv) Then, we propose a phased algorithm CVPKE-PH for the kernelized CVaR bandit problem with finitely many arms and discrete distributions, prove that it is a nearly optimal algorithm in the case of Matérn kernels (§7).

(v) Finally, we empirically compare CVPKE-UCB with a baseline algorithm in synthetic environments and demonstrate that the proposed method significantly outperforms the baseline in many settings (§8).

We shall provide all the proofs omitted in the main paper in the appendix.

2. Related Work

First, we review classical results regarding kernelized bandit problems in the conventional setting. Srinivas et al. (2010) studied the kernelized bandit problem, propose a UCB-type algorithm called GP-UCB, and provide sublinear regret upper bounds. Later, Chowdhury & Gopalan (2017) theoretically improved (Srinivas et al., 2010) by proving a self-normalized concentration inequality and provided a UCB-type algorithm called IGP-UCB. Scarlett et al. (2017) proved algorithm-independent lower bounds for the kernelized bandit problem in the case of squared exponential kernels and Matérn kernels. Although IGP-UCB is a nearly optimal algorithm for squared exponential kernels, it is not nearly optimal for Matérn kernels. A nearly optimal algorithm was provided by Valko et al. (2013) in the finitely many arms setting. More over, this has been recently pursued with more practical algorithms Camilleri et al. (2021); Salgia et al. (2021); Li & Scarlett (2022).

Next, we review related work on the CVaR or MV optimization in the multi-armed bandit or kernelized bandit literatures. If the arm set is finite and discrete and there is no model for the environment (i.e., the classical multi-armed bandit setting), the CVaR (or MV) bandit problem has been extensively studied (Tamkin et al., 2019; Baudry et al., 2021; Chang & Tan, 2022; Sani et al., 2012; Vakili & Zhao, 2016). Notably, Baudry et al. (2021) considered the CVaR bandit problem, proposed Thompson sampling strategies and proved asymptotic optimality of problem-dependent regret bounds. On the other hand, only a few results (Cakmak et al., 2020; Nguyen et al., 2021a) are known for the kernelized CVaR bandit problem. As explained in the introduction, Nguyen et al. (2021a) considered a kernelized CVaR bandit problem using the EVF model, but their method has the aforementioned drawbacks. Iwazaki et al. (2021) considered MV optimization problems with the EVF model, and their method has similar drawbacks to (Nguyen et al., 2021a). Recently, Makarova et al. (2021) also considered MV optimization problem in a kernelized bandit setting and provided a regret upper bound. Their algorithm has a parameter $m \in \mathbb{Z}_{>0}$ (k in their notation) and for each selected point x_t , observations are repeatedly sampled (*m* times) at x_t to estimate the variance of observations at x_t . Although the variance of the estimation reduces as m increase, there is a linear factor m in their regret upper bound. In this paper, using the PKE model, we can estimate MV without repeatedly sampling observations at the same point and provide a UCB-type algorithm for MV.

KME and CME were utilized in some kernelized bandit literatures (Kirschner et al., 2020; Chowdhury et al., 2020). Kirschner et al. (2020) studied a distributionally robust optimization problem with the EVF model. They used KME and maximum mean discrepancy (MMD) to define distance between two distributions. Originally, CME was used to model dynamical systems (e.g., transition probability in a Markov decision process) (Song et al., 2009; 2010). Chowdhury et al. (2020) modeled uncontrollable input state transition using CME to optimize the conditional mean of an unknown function in a kernelized bandit setting. However, they did not consider an optimization problem of risk-averse metrics as in this study.

3. Problem Formulation

In this section, we briefly review RKHS and KME, formulate our environment model called the PKE model, and provide definitions of the cumulative regret and assumptions.

RKHS and KME

First, we briefly review RKHS and KME to explain our problem formulation. For more detailed introduction to the subjects, we refer to Wendland (2004, Chapter 10), Muandet et al. (2017).

Let \mathcal{X} be a set and $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a symmetric semi-positive definite kernel, i.e., k(x, x') = k(x', x)for any x, x' and for any $x_1, \ldots, x_n \in \mathcal{X}$, matrix $(k(x_i, x_j))_{1 \le i,j \le n}$ is a semi-positive definite matrix. Let $\mathcal{F}(\mathcal{X})$ be the real vector space consisting of \mathbb{R} -valued functions on \mathcal{X} . For $x' \in \mathcal{X}$, we define a "feature vector" $\phi_k(x') \in \mathcal{F}(\mathcal{X})$ by $k(\cdot, x')$. A real Hilbert space $(\mathcal{H}_k(\mathcal{X}), \langle \cdot, \cdot \rangle_k)$ with $\mathcal{H}_k(\mathcal{X}) \subseteq \mathcal{F}(\mathcal{X})$ is called the Reproducing Kernel Hilbert Space (RKHS) associated to k if the following two properties are satisfied: (i) $\phi_k(x') \in \mathcal{H}_k(\mathcal{X})$ for all $x' \in \mathcal{X}$. (ii) $\langle f, \phi_k(x') \rangle_k = f(x')$ for all $f \in \mathcal{H}_k(\mathcal{X})$ and $x' \in \mathcal{X}$. The second property is called reproducing property, and the map $\phi_k : \mathcal{X} \to \mathcal{H}_k(\mathcal{X})$ is called the feature map. A RKHS uniquely exists if k is symmetric positive definite by the Moore-Aronszajn theorem. We provide some examples of non-linear kernels defined on $\mathbb{R}^d \times \mathbb{R}^d$. Squared Exponential (SE) kernels are defined as $k_{SE}(x,y) = \exp(-\|x-y\|_2^2/a^2)$, where a > 0 is a length scale parameter. Rational Quadratic (RQ) kernels are defined by $k_{\rm RQ}(x,y) = (1 + ||x - y||_2^2/(2\nu a^2))^{-\nu}$, where $\nu > 0$ is a parameter. Matérn- ν kernels are defined as $k_{\text{Matern}}^{(\nu)}(x,y) := \frac{2^{1-\nu}}{\Gamma(\nu)} \left(s\sqrt{2\nu}\right)^{\nu} K_{\nu}(s\sqrt{2\nu})$, where $\nu > 0$ is a parameter, $s = ||x - y||_2/a$, and K_{ν} is the modified Bessel function of the second kind. We assume $k(x, x) \leq 1$ for any $x \in \mathcal{X}$ as in the previous work (Chowdhury & Gopalan, 2017). We also assume that $\mathcal{H}_k(\mathcal{X})$ is a separable Hilbert space. This is a very weak condition and can be satisfied if k is continuous and \mathcal{X} is a separable topological space (Berlinet & Thomas-Agnan, 2011, pp. 35).

Next, we review KME (Muandet et al., 2017). Let \mathcal{Y} be a measurable space and $l : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$ a symmetric, semi-positive definite kernel. For simplicity, we assume $\sup_{y \in \mathcal{Y}} l(y, y) \leq 1$. We denote by $\mathcal{M}(\mathcal{Y})$ the set of probability measures on \mathcal{Y} . Then, there exists a unique map $\mu_l : \mathcal{M}(\mathcal{Y}) \to \mathcal{H}_l(\mathcal{Y})$ satisfying

$$\langle \mu_l(\rho), f \rangle_l = \mathbf{E}_{y \sim \rho} \left[f(y) \right]$$
 (1)

for any $f \in \mathcal{H}_l(\mathcal{Y})$ and $\rho \in \mathcal{M}(\mathcal{Y})$ (Muandet et al., 2017, Lemma 3.1), (Smola et al., 2007). The unique map μ_l is called the Kernel Mean Embedding (KME). We provide some simple examples. First, we consider the case of Dirac delta $\rho = \delta_{y'}$ with $y' \in \mathcal{Y}$. By uniqueness and the reproducing property, we see that $\mu_l(\delta_{y'}) = \phi_l(y')$. Next, we consider the case when $\mathcal{Y} = (-1, 1) \subset \mathbb{R}$ and l is the linear kernel (i.e., l(y, y') = yy'). Then $\mathcal{H}_l(\mathcal{Y})$ can be identified with \mathbb{R} with the inner product defined by the multiplication and the map μ_l can be identified with the expectation, i.e., $\mu_l(\rho) = \mathbf{E}_{y \sim \rho}[y]$. In particular, μ_l is not injective in general. However, if a power function $y \mapsto y^n$ belongs to $\mathcal{H}_l(\mathcal{Y})$, then one can recover a higher moment from $\mu_l(\rho)$. Moreover, μ_l is injective for many popular kernels such as Matérn or SE kernels, i.e., they are characteristic kernels (Gretton et al., 2006; Fukumizu et al., 2007).

Formulation of the Environment Model

Hereafter, we assume $\mathcal{Y} \subset \mathbb{R}$. In the conventional kernelized bandit problem, the reward model is given as a function from \mathcal{X} to \mathcal{Y} with some smoothness defined by the RKHS (Chowdhury & Gopalan, 2017; Srinivas et al., 2010). In this paper, rather than a function from \mathcal{X} to \mathcal{Y} , we consider a map $\rho : \mathcal{X} \to \mathcal{M}(\mathcal{Y})$ to the space of probability measures on \mathcal{Y} . More precisely, ρ is a probability kernel, that is, for any measurable set $A \subseteq \mathcal{Y}, x \mapsto \mathbf{E}_{y \sim \rho(x)} [\mathbb{1}_A(y)]$ is a measurable function, where $\mathbb{1}_A$ denotes the indicator function. However, without any smoothness or continuous assumption, we cannot hope for an algorithm with a sublinear regret guarantee. For the smoothness assumption, we consider a map $\Theta : \mathcal{H}_k(\mathcal{X}) \to \mathcal{H}_l(\mathcal{Y})$ with $\mu_l \circ \rho = \Theta \circ \phi_k$, i.e., we consider the following commutative diagram:

$$\begin{array}{cccc} \mathcal{X} & \stackrel{\rho}{\longrightarrow} & \mathcal{M}(\mathcal{Y}) \\ \text{feature map } \phi_k & & \text{KME } \mu_l \\ & & \mathcal{H}_k(\mathcal{X}) & \stackrel{\Theta}{\longrightarrow} & \mathcal{H}_l(\mathcal{Y}) \end{array} \tag{2}$$

If we regard ϕ_k (resp. μ_l) as "embeddings" to $\mathcal{H}_k(\mathcal{X})$ (resp. $\mathcal{H}_l(\mathcal{Y})$), then we can suppose that the map Θ defines the smoothness of the map ρ . In this paper, we assume that Θ is bounded linear for ρ being smooth. That is, we assume that there exists a bounded linear operator $\Theta : \mathcal{H}_k(\mathcal{X}) \to \mathcal{H}_l(\mathcal{Y})$ (i.e., $\|\Theta\|_{\text{op}} := \sup_{f \in \mathcal{H}_k(\mathcal{X}), f \neq 0} \|\Theta(f)\|_l / \|f\|_k \leq B$) that makes the diagram (2) commutative.

This can be thought as a generalization of the conventional reward model of the kernelized bandit problem (Chowdhury & Gopalan, 2017). To explain this, until the end of this paragraph, we assume that $\mathcal{Y} = (-1, 1) \subset \mathbb{R}$ and l is the linear kernel. In this case, $\mathcal{H}_l(\mathcal{Y})$ can be identified with \mathbb{R} and $\mu_l \circ \rho(x) = \mathbf{E} [\rho(x)]$. On the other hand, by the Riesz representation theorem (c.f., e.g., Conway (2019)), Θ is given by an element $f \in \mathcal{H}_k(\mathcal{X})$, i.e., $\Theta(\cdot) = \langle \cdot, f \rangle_k$. Then, by the reproducing property, $\mu_l \circ \rho = \Theta \circ \phi_k$ is equivalent to $\mathbf{E} [\rho(x)] = f(x)$ for any x, which recovers the conventional reward model in this special case.

Although this model assumption is closely related to Conditional Mean Embedding (CME) (Song et al., 2009; Park & Muandet, 2021), we did not use the existing theoretical frameworks due to some undesired properties. For example, recently, Park & Muandet (2021) provided a measure theoretic definition of CME, however, the conditional probability $P_{Y|X}$ is only defined up to a measure zero set (here X, Y are random variables on \mathcal{X} and \mathcal{Y} , respectively) and $P_{Y|X}$ cannot be necessarily regarded as a map from \mathcal{X} to $\mathcal{M}(\mathcal{Y})$. To distinguish the differences in theoretical conditions, we termed our model the PKE model.

In this paper, we consider a kernelized bandit problem as follows. Let ρ be an unknown probability kernel with commutative diagram (2), and T be a time interval. For each round $t = 1, \ldots, T$, the learner selects an arm $x_t \in \mathcal{X}$ and observes $y_t \in \mathcal{Y}$ with $y_t|_{\mathcal{F}_{t-1}} \sim \rho(x_t)$. Here \mathcal{F}_{t-1} is the σ -algebra generated by $x_1, y_1, \ldots, x_{t-1}, y_{t-1}, x_t$. We provide some more measure theoretic details in the appendix (§A).

Definition of the Cumulative Regret

To evaluate leaner's performance, we define cumulative regret for the problems. Unlike the conventional setting where the best arm is defined as an arm with the maximum expectation, we define a cumulative CVaR regret and cumulative MV regret as follows. For a distribution ρ , we define conditional value at risk $\text{CVaR}_{\alpha}(\rho)$ by $\sup_{\nu \in \mathbb{R}} \left(\nu - \frac{1}{\alpha} \mathbf{E}_{y \sim \rho} \left[\psi_{\nu}(y)\right]\right)$, where $\psi_{\nu}(y) := \max(\nu - y, 0)$. If ρ is a continuous distribution, then $\text{CVaR}_{\alpha}(\rho)$ coincides with conditional expectation $\mathbf{E}_{y \sim \rho} \left[y \mid y \leq F_{\rho}^{-1}(\alpha)\right]$ given the condition that y is less than the α -quantile of ρ , where F_{ρ} is the cumulative distribution of ρ (Acerbi & Tasche, 2002). If $\alpha = 1$, then $\text{CVaR}_{\alpha}(\rho)$ is identical to the expectation $\mathbf{E}[\rho]$, and if α is small, then it puts more focus on a rare event. Then, the cumulative CVaR regret $R_{\text{CVaR},\alpha}(T)$ is defined as

$$\sum_{t=1}^{T} \left(\sup_{x \in \mathcal{X}} \operatorname{CVaR}_{\alpha}(\boldsymbol{\rho}(x)) - \operatorname{CVaR}_{\alpha}(\boldsymbol{\rho}(x_t)) \right),$$

and the cumulative MV regret $R_{\rm MV}(T)$ is defined as

$$\sum_{t=1}^{T} \left(\sup_{x \in \mathcal{X}} \mathrm{MV}(\boldsymbol{\rho}(x)) - \mathrm{MV}(\boldsymbol{\rho}(x_t)) \right).$$

Further Assumptions on \mathcal{Y} and l

To make the problem tractable within the framework, we require further assumptions. Since we consider two kinds of objectives (CVaR and MV), we provide two assumptions respectively.

Due to technical reasons which we detail in §4 and §7, we need the following finiteness condition for the CVaR optimization.

Assumption 3.1 (Assumptions for CVaR). We assume l is positive definite and \mathcal{Y} is a finite set. Since $\psi_{\nu} \in \mathcal{H}_{l}(\mathcal{Y})$ for any $\nu \in \mathcal{Y}$, we define $U := \sup_{\nu \in \mathcal{Y}} \|\psi_{\nu}\|_{l} < \infty$.

Remark 3.2. This assumption implies $\rho(x)$ is a discrete distribution whose support is included in \mathcal{Y} for $x \in \mathcal{X}$. Under this assumption, it is easy to see that $\text{CVaR}_{\alpha}(\rho(x)) =$

$$\begin{split} \sup_{\nu \in \mathcal{Y}} \left(\nu - \frac{1}{\alpha} \mathbf{E}_{y \sim \rho} \left[\psi_{\nu}(y)\right]\right). \text{ The condition } \psi_{\nu} \in \mathcal{H}_{l}(\mathcal{Y}) \\ \text{and boundedness of } U \text{ are assured since } l \text{ is positive definite} \\ \text{and } |\mathcal{Y}| < \infty. \text{ However, if we assume } l \text{ is Matérn-1/2 and} \\ \text{take an open interval } \widetilde{\mathcal{Y}} \text{ with } \mathcal{Y} \subset \widetilde{\mathcal{Y}}, \text{ then } U \text{ has the following interpretation: } U = O(\sup_{\nu \in \widetilde{\mathcal{Y}}} \sum_{a=0}^{1} \|D^{a}\psi_{\nu}\|_{L_{2}(\widetilde{\mathcal{Y}})}) \\ (\text{Wendland, 2004, Theorem 10.47, Corollary 10.48), where } D \text{ is the weak derivative, which implies that } U \text{ is bounded} \\ \text{by a constant multiple of the length of } \widetilde{\mathcal{Y}} \text{ even if } |\mathcal{Y}| \text{ is large.} \\ \text{Assumption 3.3 (Assumptions for MV). For } n \in \mathbb{Z}_{\geq 1}, \text{ we} \\ \text{define a function } \chi_{n} : \mathcal{Y} \to \mathbb{R} \text{ as } \chi_{n}(y) = y^{n}. \text{ We assume } \\ \text{that } \chi_{1} \in \mathcal{H}_{l}(\mathcal{Y}) \text{ and } \chi_{2} \in \mathcal{H}_{l}(\mathcal{Y}). \end{split}$$

Remark 3.4. For example, this assumption is satisfied if \mathcal{Y} is a bounded open interval and l is a polynomial kernel $l(y, y') = C (1 + yy')^2$ of degree 2 with a constant C > 0, or a Matérn kernel.

Although the assumptions on l and \mathcal{Y} seem restrictive and our algorithms and regret guarantees depend on l (and $\log |\mathcal{Y}|$ in the case of CVaR), but we will see that the dependence is only through constants determined by l and \mathcal{Y} , and there is no restrictive assumption on k.

Notation

For $x_1, \ldots, x_t \in \mathcal{X}$, we define $\mathcal{H}_k(\mathcal{X})$ -coefficient row vector $\Phi_k(x_{1:t})$ by $(\phi_k(x_1), \ldots, \phi_k(x_t))$ and define $\Phi_l(y_{1:t})$ similarly for $y_1, \ldots, y_t \in \mathcal{Y}$. For $x_1, \ldots, x_t; x'_1, \ldots, x'_s \in \mathcal{X}$, we define $\mathbf{k}(x_{1:t}, x'_{1:s}) \in \mathbb{R}^{t \times s}$ as $(k(x_i, x'_j))_{1 \leq i \leq t, 1 \leq j \leq s}$.

4. UCB-type Algorithm for the Kernelized CVaR Bandits

In this section, under Assumption 3.1, we propose a UCBtype algorithm for the CVaR optimization and provide a high-probability regret upper bound. As is well-known, UCB-type algorithms elegantly solve the explore-exploit dilemma in bandit problems. Our key observation is that using the PKE model, we can model multiple functions $x \mapsto \mathbf{E}_{y \sim \rho(x)} [\psi_{\nu}(y)]$ for $\nu \in \mathcal{Y}$ with a single bounded linear map $\Theta : \mathcal{H}_k(\mathcal{X}) \to \mathcal{H}_l(\mathcal{Y})$.

First, following (Song et al., 2009; Chowdhury et al., 2020), we introduce an estimation of Θ using observation history up to time step t. For observation history $(x_1, y_1), \ldots, (x_t, y_t)$, and $x \in \mathcal{X}$, we define an estimation $\widehat{\Theta}(x; x_{1:t}, y_{1:t}) = \widehat{\Theta}_t(x) \in \mathcal{H}_l(\mathcal{Y})$ of Θ by

$$\mathbf{\Phi}_{l}(y_{1:t}) \left(\mathbf{k}(x_{1:t}, x_{1:t}) + \lambda \mathbf{1}_{t} \right)^{-1} \mathbf{k}(x_{1:t}, x).$$
(3)

We introduce some standard notations for the kernelized bandit problem. For time step $1 \le t \le T$, we define the maximum information gain $\gamma_{k,t}$ as

$$\frac{1}{2} \sup_{\widetilde{x}_1,\ldots,\widetilde{x}_t \in \mathcal{X}} \log \det \left(1_t + \lambda^{-1} k(\widetilde{x}_{1:t},\widetilde{x}_{1:t}) \right),$$

where the supremum is taken over any t points $\tilde{x}_1, \ldots, \tilde{x}_t$ in \mathcal{X} . Then, we define

$$\beta_{k,t}(\delta) = B\sqrt{\lambda} + 2\sqrt{2(\gamma_{k,t} + \log(1/\delta))}$$

where $\lambda > 0$ and $\delta > 0$ are parameters of the algorithms. We define $\sigma_{k,t}(x) \in \mathbb{R}_{\geq 0}$ by

$$\begin{split} \lambda \sigma_{k,t}^2(x) &:= k(x,x) \\ &- \mathbf{k}(x_{1:t},x)^{\mathrm{T}} \left(\mathbf{k}(x_{1:t},x_{1:t}) + \lambda \mathbf{1}_t \right)^{-1} \mathbf{k}(x_{1:t},x), \end{split}$$

for $x \in \mathcal{X}$. If t = 0, we understand $\sigma_{k,0}(x) = \lambda^{-1/2} \sqrt{k(x,x)}$ and $\widehat{\Theta}_0(x) = 0$. We also write $\sigma_{k,t}(x)$ as $\sigma_k(x; x_{1:t})$ to make the dependence of x_1, \ldots, x_t explicit.

For any $f \in \mathcal{H}_l(\mathcal{Y})$, using an existing concentration inequality for function (Durand et al., 2018, Theorem 1), Chowdhury et al. (2020) proved that following inequality in the proof of their main theorem:

$$|\langle f, \Theta \circ \phi_k(x) \rangle_l - \langle f, \widehat{\Theta}_t(x) \rangle_l| \le ||f||_l \beta_{k,t}(\delta) \sigma_{k,t}(x).$$
(4)

We provide a proof in the appendix since the technical conditions are different. Since the event on which (4) holds depends on f, in this section, we take a union bound for $f \in \{\psi_{\nu} : \nu \in \mathcal{Y}\}$. Then, by (2) and the property of the KME (1), we can derive a concentration inequality $|\mathbf{E}_{y\sim\rho(x)} [\psi_{\nu}(y)] - \langle \psi_{\nu}, \widehat{\Theta}_t(x) \rangle_l| \leq U\beta_{k,t}(|\mathcal{Y}|\delta)\sigma_{k,t}(x),$ which uniformly holds on $\nu \in \mathcal{Y}$. Therefore, we can prove the following concentration inequality for CVaR.

Proposition 4.1. For observation history $(x_1, y_1), \ldots, (x_t, y_t)$ up to time step t, we define $\widehat{\text{CVaR}}_{\alpha}(x; x_{1:t}, y_{1:t}) = \widehat{\text{CVaR}}_{\alpha,t}(x)$ by

$$\sup_{\nu \in \mathcal{Y}} \left\{ \nu - \frac{1}{\alpha} \psi_{\nu}(y_{1:t}) (\mathbf{k}(x_{1:t}, x_{1:t}) + \lambda \mathbf{1}_t)^{-1} \mathbf{k}(x_{1:t}, x) \right\},\tag{5}$$

where $\psi_{\nu}(y_{1:t})$ is defined as $(\psi_{\nu}(y_1), \dots, \psi_{\nu}(y_t))$. Then, with probability at least $1 - \delta$, the following inequality holds:

$$\left|\operatorname{CVaR}_{\alpha}(\boldsymbol{\rho}(x)) - \widehat{\operatorname{CVaR}}_{\alpha,t}(x)\right| \leq \frac{U}{\alpha} \beta_{k,t}^{(\operatorname{CV})}(\delta) \sigma_{k,t}(x),$$

for all x and t. Here $\beta_{k,t}^{(CV)}(\delta)$ is defined as $\beta_{k,t}(\delta/|\mathcal{Y}|)$, *i.e.*,

$$\beta_{k,t}^{(\mathrm{CV})}(\delta) = B\sqrt{\lambda} + 2\sqrt{2\left(\gamma_{k,t} + \log(|\mathcal{Y}|/\delta)\right)}.$$

Remark 4.2. Since ψ_{ν} is piece-wise linear, the function $\nu \mapsto \nu - \frac{1}{\alpha} \psi_{\nu}(y_{1:t}) (\mathbf{k}(x_{1:t}, x_{1:t}) + \lambda \mathbf{1}_t)^{-1} \mathbf{k}(x_{1:t}, x)$ achieves its maximum on $\{y_1, \ldots, y_t\} \cup \{\sup \mathcal{Y}, \inf \mathcal{Y}\}$. Hence, if we use Schur complement to update the matrix $(\mathbf{k}(x_{1:t}, x_{1:t}) + \lambda \mathbf{1}_t)^{-1}$ (c.f., (Chowdhury & Gopalan, 2017, Section F)), the computational complexity for $\widehat{\text{CVaR}}_{\alpha,t}(x)$ is $O(t^2)$ for each arm x.

Algorithm 1 UCB-type Algorithm for Kernelized CVaR Bandits

Input: $U > 0, \lambda > 0, \delta > 0.$ for t = 1, 2, ..., T do Define x_t as the arm with the largest UCB $\operatorname{argmax}_{x \in \mathcal{X}} \widehat{\operatorname{CVaR}}_{\alpha,t-1}(x) + \frac{U}{\alpha} \beta_{k,t-1}^{(\mathrm{CV})}(\delta) \sigma_{k,t-1}(x).$ Play x_t and observe y_t . end for

Using the confidence bounds of $\text{CVaR}(\rho(x))$ provided above, we can construct a standard UCB-type algorithm (Algorithm 1) and call it CVPKE-UCB. Compared to the existing method (Nguyen et al., 2021a) with the EVF model, our algorithm selects only the input variable x_t , and it is unnecessary to design or know the environment random variable.

Theorem 4.3. Assuming $\lambda \geq 1$, Algorithm 1 achieves the following high probability regret upper bound, i.e., for any $\delta \in (0, 1)$, the following inequality holds with probability at least $1 - \delta$:

$$R_{\text{CVaR},\alpha}(T) \le \frac{2U\sqrt{2}}{\alpha\sqrt{\log(2)}}\beta_{k,T}^{(\text{CV})}(\delta)\sqrt{T\gamma_{k,T}}.$$

Remark 4.4. (i) In the EVF-model based approach (Nguyen et al., 2021a), the kernel is defined for a pair (x, W) of the input variable x and the environment variable W. Therefore, their algorithm can incur a larger regret due to possible high dimensionality of W.

(ii) Although we assume that distribution $\rho(x)$ is discrete and \mathcal{Y} is finite, if we discretize supports of continuous distributions with \mathcal{Y} , then we can apply CVPKE-UCB to continuous distributions. In fact, we empirically show that CVPKE-UCB works well for continuous distributions in §8.

(iii) Ignoring constants α , U, $\log |\mathcal{Y}|$, CVPKE-UCB achieves the same regret upper bound as the algorithms in the conventional setting such as GP-UCB, IGP-UCB (Srinivas et al., 2010; Chowdhury & Gopalan, 2017).

(iv) By Remark 4.2, the computational complexity for the UCB for each arm at round t is given as $O(t^2)$. Therefore, CVPKE-UCB has the same computational complexity as GP-UCB and IGP-UCB. It would be possible to construct a provably efficient algorithm using kernel approximation methods as in the conventional setting (Mutnỳ & Krause, 2019; Calandriello et al., 2020; Takemori & Sato, 2021). However, for simplicity, we do not focus on efficiency in this study.

For reader's reference, we provide some known results on upper bounds of the maximum information gain $\gamma_{k,T}$. Until the end of this paragraph, we assume \mathcal{X} is a compact subset

of \mathbb{R}^d and consider *d*-dimensional linear kernels, SE kernels, RQ kernels, and Matérn- ν kernels. If *k* is a linear kernel, then $\gamma_{k,T} = O(d\log(T))$ and if *k* is an SE kernel, then $\gamma_{k,T} = O(\log^{d+1}(T))$ (Srinivas et al., 2010). If *k* is Matérn- ν , then $\gamma_{k,T} = \widetilde{O}(T^{\frac{d}{2\nu+d}})$ (Vakili et al., 2021). If *k* is a RQ kernel, then $\gamma_{k,T} = O(\log^{d+1}(T))$. This follows from (Vakili et al., 2021), (Wendland, 2004, Theorem 11.22) and, (Santin & Schaback, 2016, Theorem 15).

5. UCB-type Algorithm for Kernelized MV Bandits

To show the PKE model is sufficiently general to handle metrics other than CVaR, in this section, we consider the MV optimization problem in the kernelized setting. Under Assumption 3.3, we provide a UCB-type algorithm for minimizing the cumulative MV regret, which does not require repeated sampling unlike a recent algorithm (Makarova et al., 2021) for the same objective.

First, we provide confidence bounds for $MV(\rho)$. Recall that for $n = 1, 2, \chi_n$ denotes the function $y \mapsto y^n$. For a distribution, the *n*th raw-moment $m_n(\rho)$ is defined as $m_n(\rho) = \mathbf{E}_{y \sim \rho} [y^n]$. Then, $MV(\rho)$ is rewritten as $MV(\rho) = m_1(\rho) - cm_2(\rho) + c (m_1(\rho))^2$. Similarly to Proposition 4.1, we have the following concentration inequality for $m_n(\rho(x))$, i.e., the following inequality holds with probability at least $1 - \delta$:

$$\left| m_n(\boldsymbol{\rho}(x)) - \langle \widehat{\Theta}_t(x), \chi_n \rangle \right| \le \| \chi_n \|_l \beta_{k,t}^{(\mathrm{MV})}(\delta) \sigma_{k,t}(x),$$
(6)

for any x, t and n = 1, 2. Here $\beta_{k,t}^{(MV)}(\delta)$ is defined as $\beta_{k,t}(2\delta)$. Here by the reproducing property, the inner product $\langle \widehat{\Theta}_t(x), \chi_n \rangle$ is given as

$$(y_1^n,\ldots,y_t^n)(\mathbf{k}(x_{1:t},x_{1:t})+\lambda \mathbf{1}_t)^{-1}\mathbf{k}(x_{1:t},x).$$

Using the concentration inequality, we can easily prove the following proposition:

Proposition 5.1. For $x \in \mathcal{X}$ and observation history up to time step t, we define $\widehat{MV}_t(x)$ by

$$\widehat{\mathrm{MV}}_t(x) = \langle \widehat{\Theta}_t(x), \chi_1 \rangle - c \langle \widehat{\Theta}_t(x), \chi_2 \rangle + c \langle \widehat{\Theta}_t(x), \chi_1 \rangle^2$$

Then, with probability at least $1 - \delta$, we have the following inequality:

$$\begin{aligned} \left| \mathrm{MV}\left(\boldsymbol{\rho}(x)\right) - \widehat{\mathrm{MV}}_{t}(x) \right| \\ &\leq \mathfrak{C}\beta_{k,t}^{(\mathrm{MV})}(\delta)\sigma_{k,t}(x) + \mathfrak{C}'\left(\beta_{k,t}^{(\mathrm{MV})}(\delta)\right)^{2}\sigma_{k,t}^{2}(x), \end{aligned}$$

where $\mathfrak{C} = \|\chi_1\|_l + c\|\chi_2\|_l + 2cB\|\chi_1\|_l^2$ and $\mathfrak{C}' = c\|\chi_1\|_l^2$.

We define $\operatorname{UCB}_{t,\mathrm{MV}}(x)$ by $\widehat{\mathrm{MV}}_t(x) + \mathfrak{C}\beta_{k,t}^{(\mathrm{MV})}(\delta)\sigma_{k,t}(x) + \mathfrak{C}'(\beta_{k,t}^{(\mathrm{MV})}(\delta))^2\sigma_{k,t}^2(x)$. Then, by this proposition

UCB_{t,MV}(x) is a high-probability upper bound of MV($\rho(x)$). Therefore, we consider a UCB-type algorithm for MV, i.e., in each round t = 1, ..., T, it selects an arm x with the largest UCB_{t,MV}(x). We call the algorithm MVPKE-UCB and the following theorem provides regret upper bound of it:

Theorem 5.2. Assume $\lambda \geq 1$. Then, with probability at least $1 - \delta$, cumulative regret $R_{MV}(T)$ of MVPKE-UCB is upper bounded by

$$O\left((1+c)\beta_{k,T}^{(\mathrm{MV})}(\delta)\sqrt{T\gamma_{k,T}}+c\left(\beta_{k,T}^{(\mathrm{MV})}(\delta)\right)^{2}\gamma_{k,T}\right).$$

Remark 5.3. Recently, (Makarova et al., 2021) also proved a sublinear regret bound for the cumulative MV regret. However, we cannot directly compare our results to theirs since theoretical assumptions are different. We note that ignoring constants our regret upper bound is equivalent to that of IGP-UCB (Chowdhury & Gopalan, 2017). Ignoring constants B, λ, c , the upper bound in Theorem 5.2 is given as $O(\gamma_{k,T}\sqrt{T} + \gamma_{k,T}^2)$. On the other hand the regret upper bound of IGP-UCB is given as $O(\gamma_{k,T}\sqrt{T})$. Assuming that this upper bound is sublinear, i.e., $\gamma_{k,T} = o(\sqrt{T})$, the second term $\gamma_{k,T}^2$ in our upper bound is dominated by the first term $\gamma_{k,T}\sqrt{T}$. Therefore, regret upper bound in the theorem is equivalent to that of IGP-UCB.

In the appendix (§D), we provide additional theoretical and experimental results regarding cumulative MV regret minimization.

6. Lower Bounds of Cumulative CVaR Regret

To evaluate the difficulty of a bandit problem, it is necessary to give regret lower bounds of the problem. This section provides lower bounds for cumulative CVaR regret in the case of Matérn kernels. We focus on Matérn kernels since the corresponding RKHS has a good mathematical interpretation (i.e., Sobolev space) and Matérn- ν converges to an SE kernel if $\nu \rightarrow \infty$ (Rasmussen, 2003, section 4.2.1).

For any bandit algorithm π and an environment (ρ, Θ) , we denote by $\mathbf{E} [R_{\text{CVaR},\alpha}(T; \pi, \rho, \Theta)]$ the expected cumulative regret that algorithm π incurs up to time step T when the environment is (ρ, Θ) . Here, the expectation is taken with respect to randomness of the outputs and decisions made by the algorithm. Then, we have the following theorem.

Theorem 6.1. Assume $0 < \alpha < 1$, \mathcal{X} is the *d*dimensional cube $[0,1]^d$, and *k* is Mattern- ν kernel. Then there exist positive constants $B = B(l,\alpha)^1 T_0 =$ $T_0(\alpha, k, l), C = C(\alpha, k, l)$ satisfying the following statement. For any $T > T_0$, there exist environments

¹We write $B = B(l, \alpha)$ to indicate that constant B depends only on l and α .

 $\begin{array}{l} (\boldsymbol{\rho}_{1}, \Theta_{1}), \ldots, (\boldsymbol{\rho}_{M}, \Theta_{M}) \text{ with } \|\Theta_{m}\|_{\mathrm{op}} \leq B \mbox{ (for } 1 \leq m \leq M), \mbox{ where } M = M(\alpha, k, d, T), \mbox{ such that for any algorithm } \pi, \mbox{ we have } \mathbf{E} \left[R_{\mathrm{CVaR},\alpha}(T;\pi) \right] > CT^{\frac{\nu+d}{2\nu+d}}. \mbox{ Here } \mathbf{E} \left[R_{\mathrm{CVaR},\alpha}(T;\pi) \right] \mbox{ denotes the average expected cumulative regret } \frac{1}{M} \sum_{m=1}^{M} \mathbf{E} \left[R_{\mathrm{CVaR},\alpha}(T;\pi, \boldsymbol{\rho}_{m}, \Theta_{m}) \right]. \end{array}$

Remark 6.2. (i) This theorem can be proved in a similar manner to the conventional setting (Scarlett et al., 2017). However, unlike in the conventional setting, we have to construct probability kernels and bounded operators rather than functions. We briefly provide a sketch of the proof. We divide the cube \mathcal{X} into M small cubes with width $\Theta(T^{1/(d+2\nu)})$, where $M = \Theta(T^{d/(d+2\nu)})$. For each $m = 1, \dots, M$, we define a Bernoulli distribution $y \sim \rho_m(x)$ by y = 0 with probability $p_m(x) = \alpha(1 - f_m(x))$ and 1 otherwise. Here, f_m is a uniformly bounded function $|f_m(x)| \leq 2\varepsilon$, and is "almost zero" except for the *m*th small cube. Since $\text{CVaR}_{\alpha}(\boldsymbol{\rho}_m(x)) = \max(f_m(x), 0)$, any algorithm incurs linear regret $\Omega(\varepsilon T)$ in average. In the Matérn kernel case, it is enough to take $\varepsilon = \Theta(T^{-\nu/(d+2\nu)})$ to bound $\|\Theta_m\|_{op}$ by a constant. Therefore, we have the lower bound stated in Theorem 6.1.

(ii) If k is Matérn, then by Theorem 4.3 and $\gamma_{k,T} = \widetilde{\Theta}(T^{d/(2\nu+d)})$ (c.f., Vakili et al. (2021); Scarlett et al. (2017)), we see that CVPKE-UCB is not nearly optimal. However, the same situation occurs in the well-studied, conventional setting, i.e., IGP-UCB is not nearly optimal in the case of Matérn kernels. On the other hand, we conjecture that CVPKE-UCB is nearly optimal in the case of SE and RQ kernels.

7. Nearly Optimal Algorithm for Kernelized CVaR Bandits with Finitely Many Arms

In the last section, we provided lower bounds for cumulative CVaR regret, and remarked that CVPKE-UCB is not nearly optimal for Matérn kernels. This section provides a nearly optimal algorithm for the kernelized CVaR bandit problem. In the conventional setting, Valko et al. (2013) provided a nearly optimal algorithm for SE and Matérn kernels in the case when the arm set \mathcal{X} is finite. The concentration inequality proved by (Chowdhury & Gopalan, 2017) holds uniformly on arm x and time step t, i.e., we can take an event independent of x and t, on which the inequality holds. On the other hand, Valko et al. (2013) used the classical Azuma-Hoeffding inequality and took a union bound to obtain a concentration inequality which holds uniformly on \mathcal{X} . To the best of our knowledge, even in the wellstudied conventional setting, one has to assume \mathcal{X} is finite to achieve a nearly optimal algorithm. Therefore, in this section we assume \mathcal{X} is finite (e.g., it is a discretization of a bounded domain of a Euclidean space). Therefore, besides Assumption 3.1, we assume \mathcal{X} is a finite set in this section. Algorithm 2 Phased Algorithm for Kernelized CVaR Bandits with Finitely Many Arms and Discrete Distributions

Set
$$\mathcal{X}_{1} := \mathcal{X}, J := \lfloor \log_{2}(T) \rfloor, t_{j} = 2^{j-1}$$
 for $1 \leq j \leq J$.
Set $\widetilde{\beta}_{k,T}(\delta) := B\sqrt{\lambda} + 2\sqrt{2\log(2|\mathcal{X}||\mathcal{Y}|J/\delta)}$.
for $j = 1, \dots, J$ do
for $t = t_{j}, \dots, \min(t_{j+1} - 1, T)$ do
Play $x_{t} := \operatorname{argmax}_{x \in \mathcal{X}_{j}} \sigma_{t-1}^{(j)}(x)$ and observe y_{t} .
end for
 $s := \min(t_{j+1} - 1, T)$
 $b := \max_{x \in A_{j}} \left(\widehat{\operatorname{CVaR}}_{s}^{(j)}(x) - \frac{U}{\alpha} \widetilde{\beta}_{k,T}(\delta) \sigma_{s}^{(j)}(x) \right)$.
 $\mathcal{X}_{j+1} := \{x \in \mathcal{X}_{j} : \widehat{\operatorname{CVaR}}_{s}^{(j)}(x) + \frac{U}{\alpha} \widetilde{\beta}_{k,T}(\delta) \sigma_{s}^{(j)}(x) \geq b\}$.
end for

To apply the Azuma-Hoeffding inequality in kernelized or linear bandit settings, the selected arms should be deterministically chosen without the knowledge of the observations (Lattimore & Szepesvári, 2020, chapter 20). However, such an assumption rarely holds for meaningful bandit algorithms. To address the issue, existing methods divide time interval into phases and select arms in the phase using only observations in the previous phase (Chu et al., 2011; Valko et al., 2013). Inspired by methods for spectral bandits or misspecified linear bandits (Valko et al., 2014; Lattimore et al., 2020), we consider an algorithm that divides the time interval into J phases and call it CVPKE-PH (Algorithm 2). Here $J = \lfloor \log_2 T \rfloor$ and *j*th phase starts from t_j th round, where $t_j = 2^{j-1}$.

At the end of each phase, CVPKE-PH eliminates arms that are supposed to be suboptimal using a confidence bound. The following concentration inequality provides a tighter bound than Proposition 4.1 in this special case.

Proposition 7.1. For each phase j = 1, ..., J and $t = t_j, ..., s_j$, we define $\sigma_t^{(j)}(x)$ as $\sigma_{k,t}(x; x_{t_j:t-1})$, where $s_j := \min(t_{j+1} - 1, T)$. Similarly, we define $\widehat{\text{CVaR}}_{s_j}^{(j)}(x)$ an estimation of $\text{CVaR}(\rho(x))$ by (5) but using only the observation history $\{(x_{t_j}, y_{t_j}), ..., (x_{s_j}, y_{s_j})\}$ in the *j*th phase. Then, for any $\delta \in (0, 1)$, we have the following inequality that holds uniformly for $x \in \mathcal{X}$ and $1 \le j \le J$:

$$\left| \operatorname{CVaR}(x) - \widehat{\operatorname{CVaR}}_{s_j}^{(j)}(x) \right| \le \frac{U}{\alpha} \widetilde{\beta}_{k,T}\left(\delta\right) \sigma_{s_j}^{(j)}(x),$$

where
$$\widetilde{\beta}_{k,T}(\delta) = B\sqrt{\lambda} + 2\sqrt{2\log(2|\mathcal{X}||\mathcal{Y}|\log_2(T)/\delta)}$$
.

Using Proposition 7.1, we obtain the following result:

Theorem 7.2. We put $C := \sup_{x \in \mathcal{X}} \operatorname{CVaR}(\rho(x)) < \infty$ and assume $\lambda \ge 1$. Then, Algorithm 2 achieves the following high probability regret bound, i.e., the following



Figure 1. Cumulative CVaR Regret for Normal Environments



Figure 2. Cumulative CVaR Regret for LogNormal Environments

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

$$R_{\text{CVaR},\alpha}(T) \le 2C + \frac{8U}{\alpha \log 2} \widetilde{\beta}_{k,T}(\delta) \sqrt{T\gamma_{k,T} \log T}.$$

By combining this theorem with Theorem 6.1, we see that CVPKE-PH is a nearly optimal in the case of Matérn kernels.

Corollary 7.3. Assume that k is a Matérn kernel. Then, CVPKE-PH is a nearly optimal algorithm for the kernelized CVaR bandit problem. More precisely, the lower bound provided in Theorem 6.1 and the upper bound provided in Theorem 7.2 matches up to a polylogarithmic factor of T.

8. Experiments

In this section, we empirically evaluate our proposed algorithm in synthetic environments². Since there is no existing algorithm for the same problem formulation, we compare CVPKE-UCB (Algorithm 1) with IGP-UCB (Chowdhury & Gopalan, 2017) in two kinds (symmetric and skew output distributions) of synthetic of environments for $\alpha \in \{0.05, 0.1, 0.5, 0.9.0.95\}$. We demonstrate that the proposed method significantly outperforms the baseline if α is not close to 1.

We explain the experimental settings more in detail. In this section, we assume that \mathcal{X} is a discretization of the cube $[0,1]^d$ with d = 3, $\mathcal{X} = \{i/10 : i = 0, 1, ..., 10\}^3$, and k is Matérn-5/2 kernel with a length scale a = 0.5. We

consider two kinds of environments: symmetric and skew distributions. For the first kind of environments, we consider a family of normal distributions $\rho(x) = \mathcal{N}(\mu_m(x), \sigma_m(x))$ for $1 \le m \le 10$. Here, we constructed functions $\mu_m, \sigma_m \in$ $\mathcal{H}_k(\mathcal{X})$ independently randomly for $1 \leq m \leq 10$ (we provide details in the appendix). We note that unlike our theoretical assumptions, distribution $\rho(x)$ is continuous. We can compute the CVaR of normal distribution $\rho = \mathcal{N}(\mu, \sigma)$ explicitly. $\text{CVaR}_{\alpha}(\rho)$ is given as $\mu - \frac{\sigma}{\alpha\sqrt{2\pi}} \exp(-q^2/2)$, where $q = \Phi_{\rm SN}^{-1}(\alpha)$ and $\Phi_{\rm SN}$ is the CDF of the standard normal distribution. Since CVaR of a normal distribution $\mathcal{N}(\mu, \sigma)$ can be written as a linear combination of μ and σ , we consider more complicated environments (skew distributions). For the second kind of environments, we consider log-normal distributions $\mathcal{LN}(\mu'_m(x), \sigma'_m(x))$, where $1 \le m \le 10$ and random functions μ'_m, σ'_m are defined similarly to μ_m, σ_m . We note that the CVaR of $\rho = \mathcal{LN}(\mu, \sigma)$ is given as $\text{CVaR}_{\alpha}(\rho) = \exp(\mu + \sigma^2/2)\Phi_{\text{SN}}(q - \sigma)/\alpha$, where $q = \Phi_{SN}^{-1}(\alpha)$.

We take $\lambda = 1$ for both the algorithms, $\delta = 10^{-2}$ for IGP-UCB. To apply CVPKE-UCB to continuous distributions, assuming there exists a discretization \mathcal{Y} for support of $\rho(x)$, we let $\delta/|\mathcal{Y}| = 10^{-2}$ in the definition of $\beta_{k,t}^{(CV)}(\delta)$. Here, we note that the only dependence of $|\mathcal{Y}|$ on CVPKE-UCB is the log $|\mathcal{Y}|$ term in the definition of $\beta_{k,t}^{(CV)}(\delta)$. To tune parameters of the algorithms, we use different but the same kind of the environments using the first 200 rounds. We provide more details in the appendix.

For each α and $m = 1, \ldots, 10$, we ran both algorithms

²Source code is publicly available as a supplementary material.

in environment $\rho_m(x) = \mathcal{N}(\mu_m(x), \sigma_m(x))$. For each α and environment, we normalize the cumulative CVaR regret so that the cumulative CVaR regret of the uniform random algorithm is equal to 1 at T = 1000 round. Then we plot the mean cumulative CVaR regret for the ten environments in Figure 1. We conducted similar experiments for the lognormal environments and the results are displayed in Figure 2.

The experimental results indicate that the proposed method CVPKE-UCB achieves sublinear cumulative regret for both kinds of environments and any α while the baseline incurs linear cumulative regret in some cases. In addition, the proposed method CVPKE-UCB significantly outperforms the baseline if α is not close to 1. However, if the environments are normal distributions and α is close to 1, i.e., if CVaR_{α} is close to the expectation, then IGP-UCB is slightly better than the proposed method.

9. Conclusion

In this paper, using an environment model called PKE, we proposed a UCB-type algorithm for CVaR that improves drawbacks of the recent method. We also provided a UCBtype algorithm for MV, lower bounds for the cumulative CVaR regret, and a nearly optimal algorithm for CVaR optimization. In the experimental section, we empirically verified our theoretical result in synthetic environments.

References

- Acerbi, C. and Tasche, D. On the coherence of expected shortfall. *Journal of Banking & Finance*, 26(7):1487– 1503, 2002.
- Ash, R. B. Real analysis and probability: probability and mathematical statistics: a series of monographs and textbooks. Academic press, 2014.
- Baudry, D., Gautron, R., Kaufmann, E., and Maillard, O. Optimal Thompson Sampling strategies for supportaware CVaR bandits. In *International Conference on Machine Learning*, pp. 716–726. PMLR, 2021.
- Berlinet, A. and Thomas-Agnan, C. *Reproducing kernel Hilbert spaces in probability and statistics*. Springer Science & Business Media, 2011.
- Cakmak, S., Astudillo, R., Frazier, P. I., and Zhou, E. Bayesian Optimization of Risk Measures. In *Advances in Neural Information Processing Systems*, 2020.
- Calandriello, D., Carratino, L., Lazaric, A., Valko, M., and Rosasco, L. Near-linear time Gaussian process optimization with adaptive batching and resparsification. In *International Conference on Machine Learning*, pp. 1295– 1305. PMLR, 2020.

- Camilleri, R., Jamieson, K., and Katz-Samuels, J. Highdimensional experimental design and kernel bandits. In *International Conference on Machine Learning*, pp. 1227– 1237. PMLR, 2021.
- Chang, J. Q. and Tan, V. Y. A Unifying Theory of Thompson Sampling for Continuous Risk-Averse Bandits. Association for the Advancement of Artificial Intelligence, 2022.
- Chowdhury, S. R. and Gopalan, A. On kernelized multiarmed bandits. In *International Conference on Machine Learning*, pp. 844–853. PMLR, 2017.
- Chowdhury, S. R., Oliveira, R., and Ramos, F. Active Learning of Conditional Mean Embeddings via Bayesian Optimisation. In *Conference on Uncertainty in Artificial Intelligence*, pp. 1119–1128. PMLR, 2020.
- Chu, W., Li, L., Reyzin, L., and Schapire, R. Contextual bandits with linear payoff functions. In *Proceedings* of the Fourteenth International Conference on Artificial Intelligence and Statistics, pp. 208–214. JMLR Workshop and Conference Proceedings, 2011.
- Conway, J. B. A course in functional analysis, volume 96. Springer, 2019.
- Durand, A., Maillard, O.-A., and Pineau, J. Streaming kernel regression with provably adaptive mean, variance, and regularization. *The Journal of Machine Learning Research*, 19(1):650–683, 2018.
- Frazier, P. I. A tutorial on Bayesian optimization. *arXiv* preprint arXiv:1807.02811, 2018.
- Fukumizu, K., Gretton, A., Sun, X., and Schölkopf, B. Kernel measures of conditional dependence. In *NIPS*, volume 20, pp. 489–496, 2007.
- Gretton, A., Borgwardt, K., Rasch, M., Schölkopf, B., and Smola, A. A kernel method for the two-sample-problem. *Advances in neural information processing systems*, 19: 513–520, 2006.
- Grünewälder, S., Lever, G., Baldassarre, L., Patterson, S., Gretton, A., and Pontil, M. Conditional mean embeddings as regressors. In *Proceedings of the 29th International Conference on International Conference on Machine Learning*, pp. 1803–1810, 2012.
- Iwazaki, S., Inatsu, Y., and Takeuchi, I. Mean-variance analysis in Bayesian optimization under uncertainty. In *International Conference on Artificial Intelligence and Statistics*, pp. 973–981. PMLR, 2021.
- Kirschner, J., Bogunovic, I., Jegelka, S., and Krause, A. Distributionally robust Bayesian optimization. In *International Conference on Artificial Intelligence and Statistics*, pp. 2174–2184. PMLR, 2020.

- Lattimore, T. and Szepesvári, C. *Bandit algorithms*. Cambridge University Press, 2020.
- Lattimore, T., Szepesvari, C., and Weisz, G. Learning with good feature representations in bandits and in rl with a generative model. In *International Conference on Machine Learning*, pp. 5662–5670. PMLR, 2020.
- Li, Z. and Scarlett, J. Gaussian process bandit optimization with few batches. In *International Conference on Artificial Intelligence and Statistics*, pp. 92–107. PMLR, 2022.
- Maillard, O.-A. Self-normalization techniques for streaming confident regression. *ffhal-01349727v2*, 2016.
- Makarova, A., Usmanova, I., Bogunovic, I., and Krause, A. Risk-averse Heteroscedastic Bayesian Optimization. *Advances in Neural Information Processing Systems*, 34, 2021.
- Marzat, J., Walter, E., and Piet-Lahanier, H. Worst-case global optimization of black-box functions through Kriging and relaxation. *Journal of Global Optimization*, 55 (4):707–727, 2013.
- Minh, H. Q. Some properties of Gaussian reproducing kernel Hilbert spaces and their implications for function approximation and learning theory. *Constructive Approximation*, 32(2):307–338, 2010.
- Muandet, K., Fukumizu, K., Sriperumbudur, B., and Schölkopf, B. Kernel Mean Embedding of Distributions: A Review and Beyond. *Foundations and Trends in Machine Learning*, 10(1-2):1–144, 2017.
- Mutnỳ, M. and Krause, A. Efficient high dimensional bayesian optimization with additivity and quadrature fourier features. *Advances in Neural Information Processing Systems 31*, pp. 9005–9016, 2019.
- Nguyen, Q. P., Dai, Z., Low, B. K. H., and Jaillet, P. Optimizing conditional Value-At-Risk of black-box functions. *Advances in Neural Information Processing Systems*, 34, 2021a.
- Nguyen, Q. P., Dai, Z., Low, B. K. H., and Jaillet, P. Valueat-Risk Optimization with Gaussian Processes. *International Conference on Machine Learning*, 2021b.
- Park, J. and Muandet, K. A measure-theoretic approach to kernel conditional mean embeddings. Advances in neural information processing systems, 2021.
- Pazouki, M. and Schaback, R. Bases for kernel-based spaces. Journal of Computational and Applied Mathematics, 236 (4):575–588, 2011.

- Rasmussen, C. E. Gaussian processes in machine learning. In *Summer school on machine learning*, pp. 63–71. Springer, 2003.
- Rockafellar, R. T., Uryasev, S., et al. Optimization of conditional value-at-risk. *Journal of risk*, 2:21–42, 2000.
- Salgia, S., Vakili, S., and Zhao, Q. A domain-shrinking based Bayesian optimization algorithm with orderoptimal regret performance. *Advances in Neural Information Processing Systems*, 34:28836–28847, 2021.
- Sani, A., Lazaric, A., and Munos, R. Risk-aversion in multi-armed bandits. *Advances in Neural Information Processing Systems*, 25, 2012.
- Santin, G. and Haasdonk, B. Convergence rate of the dataindependent P-greedy algorithm in kernel-based approximation. *Dolomites Research Notes on Approximation*, 10 (Special_Issue), 2017.
- Santin, G. and Schaback, R. Approximation of eigenfunctions in kernel-based spaces. Advances in Computational Mathematics, 42(4):973–993, 2016.
- Scarlett, J., Bogunovic, I., and Cevher, V. Lower bounds on regret for noisy gaussian process bandit optimization. In *Conference on Learning Theory*, pp. 1723–1742. PMLR, 2017.
- Shahriari, B., Swersky, K., Wang, Z., Adams, R. P., and De Freitas, N. Taking the human out of the loop: A review of Bayesian optimization. *Proceedings of the IEEE*, 104(1):148–175, 2015.
- Smola, A., Gretton, A., Song, L., and Schölkopf, B. A Hilbert space embedding for distributions. In *International Conference on Algorithmic Learning Theory*, pp. 13–31. Springer, 2007.
- Song, L., Huang, J., Smola, A., and Fukumizu, K. Hilbert space embeddings of conditional distributions with applications to dynamical systems. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pp. 961–968, 2009.
- Song, L., Boots, B., Siddiqi, S., Gordon, G. J., and Smola, A. Hilbert space embeddings of hidden Markov models. 2010.
- Srinivas, N., Krause, A., Kakade, S., and Seeger, M. W. Gaussian Process Optimization in the Bandit Setting: No Regret and Experimental Design. In *International Conference on Machine Learning*, 2010.
- Takemori, S. and Sato, M. Approximation Theory Based Methods for RKHS Bandits. In *International Conference* on Machine Learning, pp. 10076–10085. PMLR, 2021.

- Tamkin, A., Keramati, R., Dann, C., and Brunskill, E. Distributionally-aware exploration for cvar bandits. In *NeurIPS 2019 Workshop on Safety and Robustness on Decision Making*, 2019.
- Vakili, S. and Zhao, Q. Risk-averse multi-armed bandit problems under mean-variance measure. *IEEE Journal of Selected Topics in Signal Processing*, 10(6):1093–1111, 2016.
- Vakili, S., Khezeli, K., and Picheny, V. On information gain and regret bounds in Gaussian process bandits. In *International Conference on Artificial Intelligence and Statistics*, pp. 82–90. PMLR, 2021.
- Valko, M., Korda, N., Munos, R., Flaounas, I., and Cristianini, N. Finite-Time Analysis of Kernelised Contextual Bandits. In *Uncertainty in Artificial Intelligence*, pp. 654, 2013.
- Valko, M., Munos, R., Kveton, B., and Kocák, T. Spectral bandits for smooth graph functions. In *International Conference on Machine Learning*, pp. 46–54. PMLR, 2014.
- Wendland, H. *Scattered data approximation*, volume 17. Cambridge university press, 2004.

Appendix

In this appendix, we provide measure theoretic details of our model in §A, remarks on the concentration inequality (4) in §B, proofs omitted in the main article in §C, additional theoretical and experimental results on MV optimization in §D, details of the experimental setting in §E.

A. Measure Theoretic Details

In this section, following (Lattimore & Szepesvári, 2020, Chapter 2, 3), we provide measure theoretic details of our bandit model. The objective of this section is to give more precise meaning of the assumption $y_t | \mathcal{F}_{t-1} \sim \rho(x_t)$ and to clarify the relationship between the expectation with respect to $\rho(x_t)$ and the conditional expectation given \mathcal{F}_{t-1} .

For a measurable space (Ω, \mathcal{F}) and a measure μ on it, we write $\mu(d\omega)$ to emphasize that μ is a measure on (Ω, \mathcal{F}) , where \mathcal{F} is a σ -algebra. For simplicity, we sometimes omit σ -algebra from notation. We recall that for measurable spaces $(\Omega, \mathcal{F}), (\Omega', \mathcal{F}')$, a probability kernel $\kappa(\cdot; d\omega')$ is a map $\kappa : \Omega \to \mathcal{M}(\Omega')$ such that $\int_A \kappa(\cdot; d\omega')$ is a measurable function for any $A \in \mathcal{F}'$. We (implicitly) assumed \mathcal{X} and \mathcal{Y} are measurable spaces $(\mathcal{X}, \mathcal{F}_x)$ and $(\mathcal{Y}, \mathcal{F}_y)$. For $t = 1, \ldots, T$, we let $\mathcal{X}^{(t)}$ and $\mathcal{F}_x^{(t)} \subseteq 2^{\mathcal{X}^{(t)}}$ be copies of \mathcal{X} and \mathcal{F}_x respectively. We define $\mathcal{Y}^{(t)}$ and $\mathcal{F}_y^{(t)}$ similarly. We define a measurable space $(\Omega, \mathcal{F}_\Omega)$ by $\Omega := \prod_{t=1}^T \mathcal{X}^{(t)} \times \mathcal{Y}^{(t)}$ and $\mathcal{F}_\Omega := \bigotimes_{t=1}^T \mathcal{F}_x^{(t)} \otimes \mathcal{F}_y^{(t)}$. We use letters ξ, ζ, ω to denote elements in \mathcal{X}, \mathcal{Y} , and $\mathcal{X} \times \mathcal{Y}$ respectively. In this section, we fix a bandit algorithm and environment model defined as follows. For $t = 1, \ldots, T$, let $\kappa_x^{(t)}(\omega_1, \ldots, \omega_{t-1}; d\xi_t)$ be a probability kernel $\kappa_x^{(t)} : \prod_{s=1}^{t-1} \mathcal{X}^{(s)} \times \mathcal{Y}^{(s)} \to \mathcal{M}(\mathcal{X}^{(t)})$, where $\omega_s = (\xi_s, \zeta_s) \in \mathcal{X}^{(s)} \times \mathcal{Y}^{(s)}$ for $s = 1, \ldots, t - 1$. If t = 1, we understand $\kappa_x^{(t)}$ is a probability measure on $\mathcal{X}^{(1)}$. The probability kernel $\kappa_x^{(t)}$ determines distribution of x_t of the algorithm in round t. For $t = 1, \ldots, T$, let $\kappa_y^{(t)}(\omega_1, \ldots, \omega_{t-1}, \xi_t; d\zeta_t)$ be a probability kernel $\kappa_y^{(t)} : \left(\prod_{s=1}^{t-1} \mathcal{X}^{(s)} \times \mathcal{Y}^{(s)}\right) \times \mathcal{X}^{(t)} \to \mathcal{M}(\mathcal{Y}^{(t)})$. The probability kernel $\kappa_y^{(t)}$ determines the output distribution of y_t in each round. The following theorem is a special case of the Ionescu-Tulcea theorem, which is a generalization of the product measure theorem.

Theorem A.1 (c.f., Ash (2014) section 2.6, Lattimore & Szepesvári (2020) Theorem 3.3). Let notation be as above. Then there exists a unique probability measure μ on $(\Omega, \mathcal{F}_{\Omega})$ such that for any $C_s = A_s \times B_s$ with $A_s \in \mathcal{F}_x^{(s)}, B_s \in \mathcal{F}_y^{(s)}$ (s = 1, ..., T),

$$\mu\left(\prod_{s=1}^{T} C_{s}\right) = \int_{A_{1}} \kappa_{x}^{(1)}(d\xi_{1}) \int_{B_{1}} \kappa_{y}^{(1)}(\xi_{1}, d\zeta_{1}) \int_{A_{2}} \kappa_{x}^{(2)}(\omega_{1}; d\xi_{2}) \int_{B_{2}} \kappa_{y}^{(2)}(\omega_{1}, \xi_{2}; d\zeta_{2}) \cdots \int_{A_{T}} \kappa_{x}^{(T)}(\omega_{1}, \dots, \omega_{T-1}; d\xi_{T}) \int_{B_{T}} \kappa_{y}^{(T)}(\omega_{1}, \dots, \omega_{T-1}, \xi_{T}; d\zeta_{T}).$$

We regard Ω as a probability space by this measure μ in the theorem and define random variables x_t, y_t for $t = 1, \ldots, T$ by projections $x_t : \Omega \to \mathcal{X}^{(t)}$ and $y_t : \Omega \to \mathcal{Y}^{(t)}$. We also define σ -algebras \mathcal{F}_{t-1} by the σ -algebra generated by $x_1, y_1, \ldots, x_{t-1}, y_{t-1}$ and x_t . For $\xi \in \mathcal{X}$, we denote probability kernel $\rho(\xi)$ by $\rho(\xi; d\zeta_t)$ to emphasize that it is a measure on $\mathcal{Y}^{(t)}$. The assumption $y_t \mid \mathcal{F}_{t-1} \sim \rho(x_t)$ means that $\kappa_y^{(t)}(\omega_1, \ldots, \omega_{t-1}, \xi_t; d\zeta_t) = \rho(\xi_t; d\zeta_t)$ for any $\omega_s \in \mathcal{X}^{(s)} \times \mathcal{Y}^{(s)}$ $(1 \le s \le t-1)$ and any $\xi_t \in \mathcal{X}^{(t)}$. By Theorem A.1 and definition of the conditional expectation, we have the following result:

Corollary A.2. Assume $y_t | \mathcal{F}_{t-1} \sim \boldsymbol{\rho}(x_t)$, or equivalently $\kappa_y^{(t)}(\omega_1, \dots, \omega_{t-1}, \xi_t; d\zeta_t) = \boldsymbol{\rho}(\xi_t; d\zeta_t)$. For any measurable $f : \mathcal{Y}^{(t)} \to \mathbb{R}$,

$$\mathbf{E}\left[f(y_t) \mid \mathcal{F}_{t-1}\right] = \int_{\mathcal{Y}} f(\zeta) \boldsymbol{\rho}(x_t; d\zeta).$$
(7)

That is, the conditional probability of Ω coincides with the probability kernel. Here, the RHS is a random variable defined as $\Omega \ni \omega \mapsto \int_{\mathcal{V}} f(\zeta) \rho(x_t(\omega); d\zeta)$.

B. Remarks on the Concentration Inequality (4)

As we remarked in the main article, the event on which inequality (4) holds depends on f. However, Chowdhury et al. (2020) took the supremum for f to prove their main theorem. Thus, their proof appears to be mistaken. Although, we

conjecture that (Chowdhury et al., 2020, Theorem 1) is true under suitable assumptions, the proof would not be trivial. Therefore, we did not use the concentration inequality provided by (Chowdhury et al., 2020) and took a union bound for $f \in \{\psi_{\nu} : \nu \in \mathcal{Y}\}$ to derive Proposition 4.1.

We introduce the following proposition, which is a weaker version of the statement of (Chowdhury et al., 2020, Theorem 1). Although it is unnecessary for our main results, it has a theoretical importance, since for example, it enables to estimate MMD between two distributions $\rho(x)$ and $\rho(x')$. We provide a sketch of a proof after the proof of Lemma C.2.

Proposition B.1. Assume $|\mathcal{Y}| < \infty$. By normalizing, we also assume that $\mathcal{Y} \subset (0, 1)$. Let $(x_1, y_1), \ldots, (x_t, y_t)$ be an observation history up to time step t. Assume l is a SE, RQ, or Matérn- ν with $\nu > 1/2$. Then, there exists a constant C = C(l) > 0 depending only on l such that the following inequality holds:

$$\mathbf{P}\left(\left|\langle f, \Theta \circ \phi_k(x) \rangle_l - \langle f, \widehat{\Theta}_t(x) \rangle_l\right| \le \|f\|_l \beta'_{k,t}(\delta) \sigma_{k,t}(x), \quad \forall f \in \mathcal{H}_l(\mathcal{Y}), \forall x \in \mathcal{X}, 1 \le \forall t \le T\right) \ge 1 - \delta_{\mathcal{H}_l}(\mathcal{Y})$$

where $\beta'_{k,t}(\delta) = B\sqrt{\lambda} + C\sqrt{2(\gamma_{k,t} + \log(\frac{|\mathcal{Y}|}{\delta}))}$. In particular, with probability at least $1 - \delta$, we have

$$\left\| \Theta \circ \phi_k(x) - \widehat{\Theta}_t(x) \right\|_l \le \beta'_{k,t}(\delta) \sigma_{k,t}(x),$$

for any t and x.

C. Proofs

This section provides proofs of the results in the main article. We provide a proof of Theorem 4.3 in §C.1, that of Theorem 5.2 in §C.2, that of Theorem 6.1 in §C.4, and that of Theorem 7.2 in §C.3.

The following lemma seems well-known³. But, we provide a proof for completeness.

Lemma C.1. For any sequence $\tilde{x}_1, \ldots, \tilde{x}_t \in \mathcal{X}$, we have

$$\log\left(\det\left(1_t + \lambda^{-1}\mathbf{k}(\widetilde{x}_{1:t}, \widetilde{x}_{1:t})\right)\right) = \sum_{s=1}^t \log\left(1 + \sigma_k^2(\widetilde{x}_t; \widetilde{x}_{1:(t-1)})\right).$$

Proof. This can be proved by induction on t using Schur complement. First, we assume t = 1. Then the assertion follows from the definition of $\sigma_k(\tilde{x}_t; \tilde{x}_{1:(t-1)})$. We assume that the assertion holds for any t' with $t' \leq t$. By Schur complement, for any $t \times t$ symmetric matrix A, $t \times 1$ matrix B, a scalar matrix D, we have

$$\det \begin{pmatrix} A & B \\ B^{\mathrm{T}} & D \end{pmatrix} = \det A \, \det \left(D - B^{\mathrm{T}} A^{-1} B \right).$$

Applying this equality, we have

$$\log \det \left(1_{t+1} + \lambda^{-1} k(\widetilde{x}_{1:t+1}, \widetilde{x}_{1:t+1}) \right) - \log \det \left(1_t + \lambda^{-1} k(\widetilde{x}_{1:t}, \widetilde{x}_{1:t}) \right) = \log(\widetilde{D}).$$

Here \widetilde{D} is given as

$$\begin{split} \widetilde{D} &= 1 + \lambda^{-1} k(\widetilde{x}_{t+1}, \widetilde{x}_{t+1}) - \lambda^{-1} k(\widetilde{x}_{1:t}, x)^{\mathrm{T}} \left(1_t + \lambda^{-1} k(\widetilde{x}_{1:t}, \widetilde{x}_{1:t}) \right) \lambda^{-1} k(\widetilde{x}_{1:t}, x) \\ &= 1 + \lambda^{-1} \left(k(\widetilde{x}_{t+1}, \widetilde{x}_{t+1}) - k(\widetilde{x}_{1:t}, x)^{\mathrm{T}} \left(\lambda 1_t + k(\widetilde{x}_{1:t}, \widetilde{x}_{1:t}) \right) k(\widetilde{x}_{1:t}, \widetilde{x}_{t+1}) \right) \\ &= 1 + \sigma_k^2(\widetilde{x}_{t+1}; \widetilde{x}_{1:t}), \end{split}$$

where the last equality follows from definition. Thus, by the induction hypothesis, the assertion of the lemma holds for t + 1. This completes the proof.

 $^{^{3}}$ For example, Srinivas et al. (2010) proved a similar result using an information theoretic definition of the information gain. Here, we provide a more algebraic proof for less restrictive assumptions.

C.1. Proof of Theorem 4.3

In this subsection, we prove regret upper bound for CVPKE-UCB.

First, we prove the following Lemma:

Lemma C.2. Let $(x_1, y_1), \ldots, (x_t, y_t)$ be an observation history up to time step t. For each $f \in \mathcal{H}_l(\mathcal{Y})$, we have the following inequality:

$$\mathbf{P}\left(\left|\langle f, \Theta \circ \phi_k(x) \rangle_l - \langle f, \widehat{\Theta}_t(x) \rangle_l\right| \le \|f\|_l \beta_{k,t}(\delta) \sigma_{k,t}(x), \quad \forall x \in \mathcal{X}, 1 \le \forall t \le T\right) \ge 1 - \delta.$$

Proof. This can be proved in a similar way to (Chowdhury et al., 2020, Theorem 1). However, since the theoretical assumptions are different, we provide a proof for completeness.

Since Θ is a bounded linear operator, it has adjoint operator $\Theta^* : \mathcal{H}_l(\mathcal{Y}) \to \mathcal{H}_k(\mathcal{X})$. We put $g = \Theta^*(f) \in \mathcal{H}_k(\mathcal{X})$. Then, we have $||g||_l \le ||\Theta^*||_{\mathrm{op}} ||f||_l \le B ||f||_l$. By (2) and the property of KME (1), we have $\langle f, \Theta \circ \phi_k(x) \rangle_l = \mathbf{E}_{y \sim \boldsymbol{\rho}(x)} [f(y)]$. On the other hand, by definition of Θ^* and the reproducing property,

$$\langle f, \Theta \circ \phi_k(x) \rangle_l = \langle g, \phi_k(x) \rangle_k = g(x).$$
(8)

Thus, it follows that

$$g(x) = \mathbf{E}_{y \sim \boldsymbol{\rho}(x)} \left[f(y) \right]. \tag{9}$$

For s = 1, ..., t, we define random variables z_s and ε_s by $z_s = f(y_s)$ and $z_s = g(x_s) + \varepsilon_s$. Then, by the reproducing property, we have

$$\langle f, \Theta_t(x) \rangle = (z_1, \dots, z_t) \left(\mathbf{k}(x_{1:t}, x_{1:t}) + \lambda \mathbf{1}_t \right)^{-1} \mathbf{k}(x_{1:t}, x).$$
 (10)

Next, we show that ε_s is conditionally $2||f||_l$ -sub-Gaussian given \mathcal{F}_{s-1} for $s = 1, \ldots, t$. We refer to Durand et al. (2018) for definition of sub-Gaussian random variables. It is enough to show that $|\varepsilon_s| \leq 2||f||_l$ and $\mathbf{E}[\varepsilon_s | \mathcal{F}_{s-1}] = 0$. The first condition follows from $|\varepsilon_s| \leq |f(y_s)| + |g(x_s)| \leq |f(y_s)| + \mathbf{E}_{y \sim \rho(x_s)}[|f(y)|]$ and $|f(y)| \leq ||f||_l ||\phi_l(y)||_l \leq ||f||_l$. The last inequality holds since we assume $\sup_{y \in \mathcal{Y}} l(y, y) \leq 1$. The second condition follows from $\mathbf{E}[f(y_s) | \mathcal{F}_{s-1}] = \mathbf{E}_{y \sim \rho(x_s)}[f(y)] = g(x_s)$ by (7) and (9). Applying (Durand et al., 2018, Theorem 1) to function g and sequence $(x_1, z_1), \ldots, (x_t, z_t)$, we have

$$|g(x) - (z_1, \dots, z_t) \left(\mathbf{k}(x_{1:t}, x_{1:t}) + \lambda \mathbf{1}_t \right)^{-1} \mathbf{k}(x_{1:t}, x) | \le ||f||_l \beta_{k,t}(\delta) \sigma_{k,t}(x),$$

with probability at least $1 - \delta$, uniformly on x and t. By (8) and (10), we have our assertion.

Although we do not require Proposition B.1 for our main results, we provide a sketch of the proof here.

Sketch of the Proof for Proposition B.1. We assume l is Matérn- ν with $\nu > 1/2$ for simplicity. The other cases can be proved similarly. Let $n = |\mathcal{Y}|$. Then, $\mathcal{H}_l(\mathcal{Y})$ is an *n*-dimensional vector space. We take a special basis of $\mathcal{H}_l(\mathcal{Y})$ called Newton basis $\{N_i\}_{i=1}^n$ with the *P*-greedy algorithm (Santin & Haasdonk, 2017; Pazouki & Schaback, 2011). This basis has the following properties:

(i) $\{N_i\}_i^n$ is an orthonormal basis of $\mathcal{H}_l(\mathcal{Y})$.

(ii) Let $m_i = \sup_{y \in \mathcal{Y}} N_i(y)$. Then, there exists a constant C' = C'(l) > 0 such that $m_i \leq C' i^{-\nu}$ for i = 1, ..., n.

Here, the second property follows from (Santin & Haasdonk, 2017, Theorem 4.1), (Pazouki & Schaback, 2011, Theorem 9.1).

For $1 \le i \le n$ and $1 \le s \le t$, let $g_i = \Theta^*(N_i) \in \mathcal{H}_k(\mathcal{X})$ and $\varepsilon_s^{(i)} = N_i(y_s) - g_i(x_s) = N_i(y_s) - \mathbf{E}_{y \sim \rho(x_s)}[N_i(y)]$. Then, we have $||g_i|| \le B$, and we see that $\varepsilon_s^{(i)}$ is $2m_i$ -sub-Gaussian by the proof of Lemma C.2. We denote by ξ_s the sth entry of the following vector

$$(\mathbf{k}(x_{1:t}, x_{1:t}) + \lambda \mathbf{1}_t)^{-1} \mathbf{k}(x_{1:t}, x)$$

for s = 1, ..., t. Define random variable Z_i by $\sum_{s=1}^t \varepsilon_s^{(i)} \xi_s$. Then by the proof of (Maillard, 2016, Corollay 5.6) or (Chowdhury & Gopalan, 2017, Theorem 2), $|Z_i| \le 2m_i \sigma_{k,t}(x) \sqrt{2(\gamma_{k,t} + \log(1/\delta))}$ with probability at least $1 - \delta$. Taking a union bound,

$$|Z_i| \le 2m_i \sigma_{k,t}(x) \sqrt{2(\gamma_{k,t} + \log(n/\delta))}$$
(11)

holds uniformly on $1 \le i \le n$, x, and t. Let $f \in \mathcal{H}_l(\mathcal{Y})$ be any element and write $f = \sum_{i=1}^n a_i N_i$. We put $g = \Theta^*(f)$ and use the same notation as in Lemma C.2. Because of orthonormality, we have $\sum_{i=1}^n a_i^2 = \|f\|_l^2$. By the proof of Lemma C.2, we note that

$$\left|g(x) - (z_1, \dots, z_t) \left(\mathbf{k}(x_{1:t}, x_{1:t}) + \lambda \mathbf{1}_t\right)^{-1} \mathbf{k}(x_{1:t}, x)\right| = \left|\langle f, \Theta \circ \phi_k(x) \rangle_l - \langle f, \widehat{\Theta}_t(x) \rangle_l\right|$$

Similarly to the proof of Lemma C.2, we apply *g* to the known concentration inequality (Chowdhury & Gopalan, 2017; Maillard, 2016). By the proof of (Maillard, 2016, Corollay 5.6) or (Chowdhury & Gopalan, 2017, Theorem 2), we have

$$|g(x) - (z_1, \dots, z_t) \left(\mathbf{k}(x_{1:t}, x_{1:t}) + \lambda \mathbf{1}_t \right)^{-1} \mathbf{k}(x_{1:t}, x)| \le B\sigma_{k,t}(x) + \sum_{s=1}^t \xi_s \varepsilon_s.$$

Since $\varepsilon_s = f(y_s) - \mathbf{E}_{y \sim \boldsymbol{\rho}(x_t)}[f(y)]$ and the linearity with respect to f, we have $\varepsilon_s = \sum_{i=1}^n a_i \varepsilon_s^{(i)}$. Therefore, $\sum_{s=1}^t \xi_s \varepsilon_s = \sum_{i=1}^n a_i \sum_{s=1}^t \xi_s \varepsilon_s^{(i)} = \sum_{i=1}^n a_i Z_i$. Thus, by inequality (11), with probability at least $1 - \delta$, we have

$$\left|\sum_{s=1}^{t} \xi_s \varepsilon_s\right| \leq \sum_{i=1}^{n} 2|a_i| m_i \sigma_{k,t}(x) \sqrt{2(\gamma_{k,t} + \log(n/\delta))}$$
$$\leq 2\sigma_{k,t}(x) \sqrt{2(\gamma_{k,t} + \log(n/\delta))} \sqrt{\sum_{i=1}^{n} a_i^2} \sqrt{\sum_{i=1}^{n} m_i^2}$$
$$\leq 2C' \|f\|_l \sigma_{k,t}(x) \sqrt{2(\gamma_{k,t} + \log(n/\delta))} \sqrt{\sum_{i=1}^{n} i^{-2\nu}}.$$

Since $\sqrt{\sum_{i=1}^{n} i^{-2\nu}} < \sqrt{\zeta(2\nu)}$, where ζ is the Riemann zeta, we have our assertion.

Next, we prove Proposition 4.1.

Proof of Proposition 4.1. By assumption $\|\psi_{\nu}\|_{l} \leq U$ uniformly on ν . Then by Lemma C.2 and taking a union bound, with probability at least $1 - \delta$, we have the following inequality:

$$\left|\langle \psi_{\nu}, \Theta \circ \phi_{k}(x) \rangle_{l} - \langle \psi_{\nu}, \widehat{\Theta}_{t}(x) \rangle_{l}\right| \leq U \beta_{k,t}^{(\mathrm{CV})}(\delta) \sigma_{k,t}(x),$$

for any $\nu \in \mathcal{Y}, x \in \mathcal{X}, t = 1, ..., T$. By the commutativity (2) and the property of KME (1), we have $\langle \psi_{\nu}, \Theta \circ \phi_k(x) \rangle_l = \mathbf{E}_{y \sim \boldsymbol{\rho}(x)} [\psi_{\nu}(y)]$. By the reproducing property, we have $\langle \psi_{\nu}, \widehat{\Theta}_t(x) \rangle_l = \psi_{\nu}(y_{1:t}) K_t^{-1} \mathbf{k}(x_{1:t}, x)$, where $K_t := \mathbf{k}(x_{1:t}, x_{1:t}) + \lambda \mathbf{1}_t$. Thus, it follows that

$$\nu - \frac{1}{\alpha} \psi_{\nu}(y_{1:t}) K_t^{-1} \mathbf{k}(x_{1:t}, x) - \frac{U}{\alpha} \beta_{k,t}^{(\mathrm{CV})}(\delta) \sigma_{k,t}(x)$$

$$\leq \nu - \frac{1}{\alpha} \mathbf{E}_{y \sim \boldsymbol{\rho}(x)} \left[\psi_{\nu}(y) \right] \leq \nu - \frac{1}{\alpha} \psi_{\nu}(y_{1:t}) K_t^{-1} \mathbf{k}(x_{1:t}, x) + \frac{U}{\alpha} \beta_{k,t}^{(\mathrm{CV})}(\delta) \sigma_{k,t}(x).$$

Taking supremum for $\nu \in \mathcal{Y}$, we obtain the assertion of the proposition.

Proof of Theorem 4.3. This can be proved by a standard argument. Let E be an event on which the inequality in Proposition 4.1 holds. Then $\mathbf{P}(E) \ge 1 - \delta$ by Proposition 4.1. We define the best arm x^* by $\operatorname{argmax}_{x \in \mathcal{X}} \operatorname{CVaR}_{\alpha}(\boldsymbol{\rho}(x))$. Then on

event E, the following inequalities hold:

$$R_{\text{CVaR},\alpha}(T) = \sum_{t=1}^{T} \{ \text{CVaR}_{\alpha}(\boldsymbol{\rho}(x^*)) - \text{CVaR}_{\alpha}(\boldsymbol{\rho}(x_t)) \}$$

$$\leq \sum_{t=1}^{T} \left\{ \widehat{\text{CVaR}}_{\alpha,t}(x^*) + \frac{U}{\alpha} \beta_{k,t-1}^{(\text{CV})}(\delta) \sigma_{k,t-1}(x^*) - \widehat{\text{CVaR}}_{\alpha,t}(x_t) + \frac{U}{\alpha} \beta_{k,t-1}^{(\text{CV})}(\delta) \sigma_{k,t-1}(x_t) \right\}$$

$$\leq \frac{2U}{\alpha} \sum_{t=1}^{T} \beta_{k,t-1}^{(\text{CV})}(\delta) \sigma_{k,t-1}(x_t).$$

Here the first inequality follows the definition of E, and the second inequality holds because CVPKE-UCB selects the arm with the largest UCB in each round. Since $\beta_{k,t-1}(\delta) \leq \beta_{k,T}(\delta)$ by Lemma C.1, we further have

$$\begin{aligned} R_{\text{CVaR},\alpha}(T) &\leq \frac{2U}{\alpha} \beta_{k,T}^{(\text{CV})}(\delta) \sum_{t=1}^{T} \sigma_{k,t-1}(x_t) \\ &\leq \frac{2U}{\alpha} \beta_{k,T}^{(\text{CV})}(\delta) \sqrt{T} \sqrt{\sum_{t=1}^{T} \sigma_{k,t-1}^2(x_t)} \\ &\leq \frac{2U}{\alpha \sqrt{\log(2)}} \beta_{k,T}^{(\text{CV})} \sqrt{T} \sqrt{\sum_{t=1}^{T} \log\left(1 + \sigma_{k,t-1}^2(x_t)\right)} \\ &\leq \frac{2U\sqrt{2}}{\alpha \sqrt{\log(2)}} \beta_{k,T}^{(\text{CV})}(\delta) \sqrt{T\gamma_{k,T}}. \end{aligned}$$

Here the second inequality follows from the Cauchy-Schwartz inequality, the third inequality follows from $\log(1+u) \ge \log(2)u$ for any $0 \le u \le 1$, and the 4th inequality follows from Lemma C.1 and definition of the maximum information gain $\gamma_{k,T}$. This completes the proof.

C.2. Proof Theorem 5.2

In this section, we prove theoretical results (Proposition 5.1 and Theorem 5.2) regarding the kernelized MV bandit problem. First, we can prove inequality (6) by Lemma C.2 and taking a union bound. In this subsection, we fix an event E on which inequality (6) holds. We can prove Proposition 5.1 using confidence bounds for raw moments (6) as follows.

Proof of Proposition 5.1. We assume the event *E* holds. By multiplying $|m_1(\rho(x))|$ of both sides of inequality (6) with m = 1, we have

$$\left| m_{1}^{2}(\boldsymbol{\rho}(x)) - \langle \widehat{\Theta}_{t}(x), \chi_{1} \rangle m_{1}(\boldsymbol{\rho}(x)) \right| \leq \left| m_{1}(\boldsymbol{\rho}(x)) \right| \|\chi_{1}\|_{l} \beta_{k,t}(\delta) \sigma_{k,t}(x) \leq B \|\chi_{1}\|_{l}^{2} \beta_{k,t}^{(\mathrm{MV})}(\delta) \sigma_{k,t}(x).$$
(12)

Here the last inequality follows from boundedness $|m_1(\rho(x))| = |\langle \Theta(\phi_k(x)), \chi_1 \rangle| \le ||\chi_1||_l B$. Again by inequality (6) with m = 1, we have

$$\left| \langle \widehat{\Theta}_t(x), \chi_1 \rangle \right| \le | \langle \Theta(\phi_k(x)), \chi_1 \rangle | + \| \chi_1 \|_l \beta_{k,t}^{(\mathrm{MV})}(\delta) \sigma_{k,t}(x) \le B \| \chi_1 \|_l + \| \chi_1 \|_l \beta_{k,t}^{(\mathrm{MV})}(\delta) \sigma_{k,t}(x).$$

By this inequality and multiplying $|\langle \widehat{\Theta}_t(x), \chi_1 \rangle|$ of both sides of inequality (6) with m = 1, we have

$$\left|\langle\widehat{\Theta}_t(x),\chi_1\rangle^2 - \langle\widehat{\Theta}_t(x),\chi_1\rangle m_1(\boldsymbol{\rho}(x))\right| \le B \|\chi_1\|_l^2 \beta_{k,t}^{(\mathrm{MV})}(\delta)\sigma_{k,t}(x) + \|\chi_1\|_l^2 \left(\beta_{k,t}^{(\mathrm{MV})}(\delta)\right)^2 \sigma_{k,t}^2(x).$$
(13)

By inequalities (12) and (13), we have

$$\left| m_1^2(\boldsymbol{\rho}(x)) - \langle \widehat{\Theta}_t(x), \chi_1 \rangle^2 \right| \le 2B \|\chi_1\|_l^2 \beta_{k,t}^{(\mathrm{MV})}(\delta) \sigma_{k,t}(x) + \|\chi_1\|_l^2 \left(\beta_{k,t}^{(\mathrm{MV})}(\delta) \right)^2 \sigma_{k,t}^2(x).$$

Then, by $MV(\rho(x)) = m_1(\rho(x)) - cm_2(\rho(x)) + c(m_1(\rho(x)))^2$ and inequality (6) with n = 1, 2, the assertion of the proposition holds on even E, where $\mathbf{P}(E) \leq 1 - \delta$ by definition.

Then, we can prove Theorem 5.2 similarly to Theorem 4.3.

Proof of Theorem 5.2. Let *E* be an event with $\mathbf{P}(E) \ge 1 - \delta$ on which the inequality in Proposition 5.1 holds. We define x^* by $\operatorname{argmax} MV(\boldsymbol{\rho}(x))$. Then, on event *E*, we have

$$\begin{split} R_{\rm MV}(T) &= \sum_{t=1}^{T} \left\{ {\rm MV}(\boldsymbol{\rho}(x^*)) - {\rm MV}(\boldsymbol{\rho}(x_t)) \right\} \\ &\leq \sum_{t=1}^{T} \left\{ \widehat{{\rm MV}}_{t-1}(x^*) + \mathfrak{C}\beta_{k,t-1}^{({\rm MV})}(\delta)\sigma_{k,t-1}(x^*) + \mathfrak{C}' \left(\beta_{k,t-1}^{({\rm MV})}(\delta)\right)^2 \sigma_{k,t-1}^2(x^*) \right. \\ &\left. - \widehat{{\rm MV}}_{t-1}(x_t) + \mathfrak{C}\beta_{k,t-1}^{({\rm MV})}(\delta)\sigma_{k,t-1}(x_t) + \mathfrak{C}' \left(\beta_{k,t-1}^{({\rm MV})}(\delta)\right)^2 \sigma_{k,t-1}^2(x_t) \right\} \\ &\leq 2\mathfrak{C}\sum_{t=1}^{T} \beta_{k,t-1}^{({\rm MV})}(\delta)\sigma_{k,t-1}(x_t) + 2\mathfrak{C}' \sum_{t=1}^{T} \left(\beta_{k,t-1}^{({\rm MV})}(\delta)\right)^2 \sigma_{k,t-1}^2(x_t). \end{split}$$

Here the first inequality follows from Proposition 5.1 and the second inequality holds since the algorithm selects an arm x_t with the largest UCB in each round. By the proof of Theorem 4.3, we see that $\beta_{k,t-1}(\delta) \leq \beta_{k,T}(\delta)$, $\sum_{t=1}^{T} \sigma_{k,t-1}(x_t) = O(\sqrt{T\gamma_{k,T}})$, and $\sum_{t=1}^{T} \sigma_{k,t-1}^2(x_t) = O(\gamma_{k,T})$. Thus, we have the assertion of the theorem.

C.3. Proof of Theorem 7.2

In this section, we prove a regret upper bound of the phased algorithm CVPKE-PH.

The following proposition is essentially due to (Valko et al., 2013), but for the completeness, we give a proof.

Proposition C.3. Let $\theta \in \mathcal{H}_k(\mathcal{X})$ with $\|\theta\|_k \leq B$ and $x_1, \ldots, x_t \in \mathcal{X}$. Assume random variables $\varepsilon_1, \ldots, \varepsilon_t$ are bounded $|\varepsilon_s| \leq R$ for $1 \leq s \leq t$ and a martingale difference with respect to a filtration $\{\mathcal{G}_s\}_{s=1}^t$, i.e., there exists a sequence $\{\mathcal{G}_s\}_{s=1}^t$ of σ -algebras such that $\mathcal{G}_1 \subseteq \mathcal{G}_2 \cdots \subseteq \mathcal{G}_t$ and ε_s is \mathcal{G}_s -measurable for each $1 \leq s \leq t$. We put $y_s = \theta(x_s) + \varepsilon_s$ for $s = 1, \ldots, t$. For $x \in \mathcal{X}$, we define $\hat{\mu}(x; x_{1:t}, y_{1:t}) = \hat{\mu}_t(x)$ by

$$\widehat{\mu}_t(x) := (y_1, \dots, y_t) \left(\mathbf{k}(x_{1:t}, x_{1:t}) + \lambda \mathbf{1}_t \right)^{-1} \mathbf{k}(x_{1:t}, x).$$

Then, for each x, we have the following inequality:

$$\mathbf{P}\left(\left|\widehat{\mu}_t(x) - \theta(x)\right| \le \left(R\sqrt{2\log(2/\delta)} + \lambda^{1/2} \|\theta\|_k\right) \sigma_{k,t}(x)\right) \ge 1 - \delta,$$

for any $\delta \in (0, 1)$.

Proof. In this proof, we sometimes use matrix notation for the inner product of RKHS as in the previous work (Valko et al., 2013), i.e., we write $\phi_k(x)^T \phi_k(x') = \langle \phi_k(x), \phi_k(x') \rangle_k$ for $x, x' \in \mathcal{X}$. In addition, since we only consider RKHS $\mathcal{H}_k(\mathcal{X})$ (and not $\mathcal{H}_l(\mathcal{Y})$) in this proof, we sometimes drop k from notation. We put $K_t := \mathbf{k}(x_{1:t}, x_{1:t}) + \lambda \mathbf{1}_t$ and define Φ_t by $(\phi(x_1), \ldots, \phi(x_t))^T$. Then, $K_t = \Phi_t \Phi_t^T + \lambda \mathbf{1}_t$. We define $C_t := \Phi_t^T \Phi_t + \lambda I$, where I is the identity map of $\mathcal{H}(\mathcal{X})$. More precisely, $C_t : \mathcal{H}(\mathcal{X}) \to \mathcal{H}(\mathcal{X})$ is given as $C_t(f) = \sum_{s=1}^t \langle f, \phi(x_s) \rangle \phi(x_s) + \lambda f$. Then, C_t is a finite-rank, self-adjoint, positive-definite operator. Since $\lambda > 0$, C_t has its inverse $C_t^{-1} : \mathcal{H}(\mathcal{X}) \to \mathcal{H}(\mathcal{X})$ and C_t^{-1} is also a self-adjoint, bounded linear operator. Since $(\Phi_t^T \Phi_t + \lambda I) \Phi_t^T = \Phi_t^T (\Phi_t \Phi_t^T + \lambda \mathbf{1}_t)$, we have $\Phi_t^T (\Phi_t \Phi_t^T + \lambda \mathbf{1}_t)^{-1} = (\Phi_t^T \Phi_t + \lambda I)^{-1} \Phi_t^T$. Therefore, for any $\mathbf{a} = (a_1, \ldots, a_t)^T \in \mathbb{R}^t$, we have

$$C_{t}^{-1}\left(\sum_{s=1}^{t}\phi(x_{s})a_{s}\right) = C_{t}^{-1}\Phi_{t}^{\mathrm{T}}\mathbf{a} = \Phi_{t}^{\mathrm{T}}K_{t}^{-1}\mathbf{a} = \mathbf{a}^{\mathrm{T}}K_{t}^{-1}\Phi_{t}.$$
(14)

By definition of $\hat{\mu}_t(x)$ and the reproducing property, we have $\hat{\mu}_t(x) = \langle y_{1:t} K_t^{-1} \Phi_t, \phi(x) \rangle$. Thus, by (14), we have

 $\widehat{\mu}_t(x) = \langle C_t^{-1} \sum_{s=1}^t \phi(x_s) y_s, \phi(x) \rangle.$ Therefore,

$$\widehat{\mu}_{t}(x) - \theta(x) = \left\langle C_{t}^{-1} \sum_{s=1}^{t} \phi(x_{s}) y_{s}, \phi(x) \right\rangle - \left\langle \theta, \phi(x) \right\rangle$$

$$= \left\langle C_{t}^{-1} \sum_{s=1}^{t} \phi(x_{s}) \left(\left\langle \theta, \phi(x_{s}) \right\rangle + \varepsilon_{s} \right), \phi(x) \right\rangle - \left\langle \theta, \phi(x) \right\rangle$$

$$= \left\langle C_{t}^{-1} (C_{t}(\theta) - \lambda \theta) + C_{t}^{-1} \sum_{s=1}^{t} \phi(x_{s}) \varepsilon_{s}, \phi(x) \right\rangle - \left\langle \theta, \phi(x) \right\rangle$$

$$= - \left\langle \lambda C_{t}^{-1} \theta, \phi(x) \right\rangle + \sum_{s=1}^{t} \left\langle C_{t}^{-1} \phi(x_{s}), \phi(x) \right\rangle \varepsilon_{s}.$$
(15)

Here the second and third equalities follows from definitions of y_s and C_t . To apply the Azuma-Hoeffding inequality, we compute $\sum_{s=1}^t \langle C_t^{-1}\phi(x_s), \phi(x) \rangle^2$. To do this, for $f, g \in \mathcal{H}(\mathcal{X})$, we define an inner product $\langle f, g \rangle_{C_t^{-1}} := \langle C_t^{-1}f, g \rangle_k$ Then, since C_t is self-adjoint and $\lambda > 0$, $\langle f, f \rangle_{C_t^{-1}} \ge 0$. Thus, we write $\|f\|_{C_t^{-1}} := \sqrt{\langle f, f \rangle_{C_t^{-1}}}$. We also define a linear operator D_t by letting $\lambda = 0$ in C_t , i.e., $D_t(f) = \sum_{s=1}^t \langle \phi(x_s), f \rangle \phi(x_s)$. We define norms $\|f\|_{D_t}$ and $\|f\|_{C_t}$ in the same way as $\|f\|_{C_t^{-1}}$. Then, for any $f \in \mathcal{H}(\mathcal{X})$, we have $\|f\|_{D_t} \le \|f\|_{C_t}$ because $\lambda > 0$. Thus, we have

$$\sum_{s=1}^{t} \langle C_t^{-1} \phi(x_s), \phi(x) \rangle^2 = \sum_{s=1}^{t} \langle \phi(x_s), C_t^{-1} \phi(x) \rangle^2 = \|C_t^{-1} \phi(x)\|_{D_t}^2 \le \|C_t^{-1} \phi(x)\|_{C_t}^2 = \|\phi(x)\|_{C_t^{-1}}^2.$$

Here, the first equality follows since C_t^{-1} is self-adjoint, the second and the third equalities follows from definition of the norms. Next, we show that $\|\phi(x)\|_{C_t^{-1}}^2 = \sigma_t^2(x)$. By defition and reproducing property, we have $C_t\phi(x) = \sum_{s=1}^t k(x, x_s)\phi(x_s) + \lambda\phi(x)$. By applying C_t^{-1} and (14), we obtain $\phi(x) = k(x, x_{1:t})K_t^{-1}\Phi_t + \lambda C_t^{-1}\phi(x)$. By this equality and defition of $\|\phi(x)\|_{C_t^{-1}}$, we see that $\|\phi(x)\|_{C_t^{-1}}^2 = \sigma_t^2(x)$ (c.f., (Valko et al., 2013)). Since $\langle C_t^{-1}\phi(x_s), \phi(x)\rangle\varepsilon_s$ is martingale difference with respect to the filtration $\{\mathcal{G}_s\}_s$, by the Azume-Hoeffding theorem, the absolute value of the second term in (15) is bounded by $R\sqrt{2\log(2/\delta)}\sigma_t(x)$ with probability at least $1 - \delta$. For the first term,

$$|\langle \lambda C_t^{-1}\theta, \phi(x) \rangle| = |\langle \lambda \theta, \phi(x) \rangle_{C_t^{-1}}| \le \lambda \|\theta\|_{C_t^{-1}} \|\phi(x)\|_{C_t^{-1}} \le \lambda^{1/2} \|\theta\|_k \sigma_t(x).$$

Here the first inequality follows from Cauch-Schwartz and the second inequality follows from the fact that all the eigenvalues of C_t^{-1} is less than or equal to λ^{-1} .

Next, we prove Proposition 7.1.

Proof of Proposition 7.1. For $t = t_j, \ldots, s_j$, let x_t (resp. y_t) be the arm selected by (resp. output observed by) CVPKE-PH in round t, where $s_j := \min(t_{j+1} - 1, T)$. Denote by $\Theta^{(j)}(x)$ the estimation $\widehat{\Theta}(x; x_{t_j:s_j}, y_{t_j:s_j})$ of $\Theta(\phi_k(x))$ using only history $\{(x_t, y_t)\}_{t_j \le t \le s_j}$ in phase j defined by (3).

Since Θ is a bounded operator, the adjoint operator $\Theta^* : \mathcal{H}_l(\mathcal{Y}) \to \mathcal{H}_k(\mathcal{X})$ exists and satisfies $\|\Theta^*\|_{\text{op}} \leq B$. We take any $f \in \mathcal{H}_l(\mathcal{Y})$ with $\|f\|_l \leq U$ and put $g = \Theta^*(f) \in \mathcal{H}_k(\mathcal{X})$. Then, we have $\|g\|_k \leq BU$. Since Θ^* is adjoint, we have $\langle f, \Theta \circ \phi_k(x) \rangle_l = \langle g, \phi_k(x) \rangle_k = g(x)$. On the other hand, we have $\langle f, \Theta \circ \phi_k(x) \rangle_l = \langle f, \nu_l(\rho(x)) = \mathbf{E}_{y \sim \rho(x)}[f(y)]$. By the reproducing property, we have

$$\langle f, \Theta^{(j)}(x) \rangle_l = (f(y_{t_j}), \dots, f(y_{s_j})) (\mathbf{k}(x_{1:t}, x_{1:t}) + \lambda \mathbf{1}_t)^{-1} \mathbf{k}(x_{1:t}, x).$$

For each $t = t_j, \ldots, s_j$, we put $\varepsilon_t = g(x_t) - f(y_t)$. Then similarly to the proof of Lemma C.2, we see that ε_t satisfies $|\varepsilon_t| \le 2||f||_l$ and $\mathbf{E}[\varepsilon_t | \mathcal{F}_{t-1}] = 0$. Thus, by Proposition C.3, for each f and x, we have

$$\left|g(x) - \langle f, \Theta^{(j)}(x) \rangle_l\right| = \left|\mathbf{E}_{y \sim \boldsymbol{\rho}(x)} \left[f(y)\right] - \langle f, \Theta^{(j)}(x) \rangle_l\right| \le U \left(B\sqrt{\lambda} + 2\sqrt{2\log(2/\delta)}\right) \sigma_{s_j}^{(j)}(x), \tag{16}$$

with probability at least $1 - \delta$. More precisely, when deriving inequality (16), using notation in section A, we fix $\omega_1, \ldots, \omega_{t_j-1}$, consider probability space $\prod_{t=t_j}^T \mathcal{X}^{(t)} \times \mathcal{Y}^{(t)}$ by Theorem A.1, apply Proposition C.3, and take expectation with respect to $\omega_1, \ldots, \omega_{t_j-1}$. The statement of the proposition follows from by taking a union bound for $f \in \{\psi_{\nu} : \nu \in \mathcal{Y}\}$, $x \in \mathcal{X}, j = 1, \ldots, J$ in (16) and by definition of CVaR.

Lemma C.4. Assume $\lambda \geq 1$. For any $x \in A_j$, we have

$$\left(\sigma_{s_j}^{(j)}(x)\right)^2 \le \frac{2\gamma_T}{(s_j - t_j + 1)\log 2}.$$

Proof. For any $x \in \mathcal{X}_i$, we have the following inequality:

$$\left(\sigma_{s_j}^{(j)}(x)\right)^2 \le \frac{1}{s_j - t_j + 1} \sum_{s=t_j}^{s_j} \left(\sigma_{s-1}^{(j)}(x)\right)^2$$

$$\le \frac{1}{s_j - t_j + 1} \sum_{s=t_j}^{s_j} \left(\sigma_{s-1}^{(j)}(x_s)\right)^2$$

$$\le \frac{1}{(s_j - t_j + 1)\log 2} \sum_{s=t_j}^{s_j} \log\left(1 + \left(\sigma_{s-1}^{(j)}(x_s)\right)^2\right)$$

Here, the first inequality follows from $\sigma_{s_j}^{(j)}(x) \leq \sigma_{s-1}^{(j)}(x)$, for any $x \in \mathcal{X}_j, t_j \leq s \leq s_j$. The second inequality follows by the choice of x_s . By the assumption on the kernel $\sup_{x \in \mathcal{X}} k(x, x) \leq 1$ and $\lambda \geq 1$, we have $\sigma^{(j)}(x_s) \in [0, 1]$. Then, the third inequality follows from $\log(1+u) \geq \log(2)u$ for any $u \in [0, 1]$. The statement of the lemma follows from Lemma C.1 and definition of γ_T .

Proof of Theorem 7.2. Until the end of this subsection, we denote by E an event on which the following inequality holds:

$$\left| \operatorname{CVaR}(x) - \widehat{\operatorname{CVaR}}_{s_j}^{(j)}(x) \right| \leq \frac{U}{\alpha} \widetilde{\beta}_T(\delta) \sigma_{s_j}^{(j)}(x), \quad \forall x \in \mathcal{X}, \quad 1 \leq j \leq J.$$

By Proposition 7.1 and taking union bounds for $x \in \mathcal{X}$ and j = 1, ..., J, we have $\mathbf{P}(E) \ge 1 - \delta$.

Let $1 \le t \le T$ and assume that t is in the j + 1th phase with $0 \le j \le J - 1$. In this proof, for simplicity, we denote $\sigma_{s_j}^{(j)}$ by $\sigma^{(j)}$, $\widehat{\text{CVaR}}_{s_j}^{(j)}$ by $\widehat{\text{CVaR}}_{s_j}^{(j)}$, and $\widetilde{\beta}_{k,T}(\delta)$ by $\widetilde{\beta}$. Put $x^* := \operatorname{argmax}_{x \in \mathcal{X}} \operatorname{CVaR}_{\alpha}(\rho(x))$. Since x^* and x_t are not eliminated in the *j*th phase, we have

$$\widehat{\text{CVaR}}^{(j)}(x_t) + \frac{U\widetilde{\beta}}{\alpha}\sigma^{(j)}(x_t) \ge \max_{x \in \mathcal{X}_j} \left(\widehat{\text{CVaR}}^{(j)}(x) - \frac{U\widetilde{\beta}}{\alpha}\sigma^{(j)}(x) \right)$$
$$\ge \widehat{\text{CVaR}}^{(j)}(x^*) - \frac{U\widetilde{\beta}}{\alpha}\sigma^{(j)}(x^*).$$
(17)

We put $r_t = \text{CVaR}_{\alpha}(x^*) - \text{CVaR}_{\alpha}(x_t)$. Then on even E, we have

$$r_{t} \leq \widehat{\text{CVaR}}^{(j)}(x^{*}) - \widehat{\text{CVaR}}^{(j)}(x_{t}) + \frac{U\widetilde{\beta}}{\alpha}(\sigma^{(j)}(x^{*}) + \sigma^{(j)}(x_{t}))$$

$$\leq \frac{2U\widetilde{\beta}}{\alpha}(\sigma^{(j)}(x^{*}) + \sigma^{(j)}(x_{t})).$$
(18)

Here inequality (18) follows from (17). On event E, we have the following inequalities:

$$\begin{aligned} R_{\text{CVaR},\alpha}(T) &= \sum_{t=1}^{T} r_t \le 2C + \sum_{t=2}^{T} r_t \\ &\le 2C + \sqrt{T-1} \sqrt{\sum_{j=1}^{J-1} \sum_{t=t_{j+1}}^{s_{j+1}} r_t^2} \\ &\le 2C + \frac{4U\widetilde{\beta}}{\alpha} \sqrt{(T-1)\gamma_T} \sqrt{\sum_{j=1}^{J-1} \sum_{t=t_{j+1}}^{s_{j+1}} \frac{2}{(s_j - t_j + 1)\log 2}} \\ &\le 2C + \frac{8U\widetilde{\beta}}{\alpha \log 2} \sqrt{(T-1)\gamma_T \log T}. \end{aligned}$$

Here, in the first inequality we decompose the sum $\sum_{t=1}^{T} r_t$ as $r_1 + \sum_{t=2}^{T} r_t$ because the previous phase does not exist for the first round t = 1. The second inequality follows from Cauchy-Schwartz, the third inequality follows from (18) and Lemma C.4. The 4th inequality follows from $J \leq \log_2 T$ and $s_{j+1} - t_{j+1} + 1 \leq 2(s_j - t_j + 1)$.

C.4. Proof of Theorem 6.1

In this section following (Scarlett et al., 2017), we provide a proof of lower bounds for the cumulative CVaR regret.

Since we construct distributions $\{\rho(x)\}_x$ by a discrete distribution, we introduce the following lemma, which can be easily proved by the definition of CVaR.

Lemma C.5. Let $p, q \ge 0$ with p + q = 1 and $a, b \in \mathcal{Y}$. We define a discrete distribution ρ by y = a with probability p and y = b with probability q. Then, we have

$$CVaR_{\alpha}(\rho) = \begin{cases} a & \text{if } \alpha \leq p \\ \frac{p}{\alpha}a + (1 - \frac{p}{\alpha})b & \text{if } p \leq \alpha \end{cases}$$

For $f \in L_1(\mathbb{R}^d)$, we define Fourier transform \hat{f} of f by

$$\hat{f}(x) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} f(y) \exp(-ix \cdot y) dy.$$

We define a compactly supported smooth function $H : \mathbb{R}^d \to \mathbb{R}$ by

$$H(x) := \begin{cases} \exp\left(-\frac{-1}{1 - \|x\|_2^2}\right) & \|x\|_2^2 \le 1\\ 0 & \|x\|_2 > 1. \end{cases}$$

We put $h = \hat{H}$. Then, by definition of H, $\sup_{x \in \mathbb{R}^d} h(x) \le h(0) > 0$. Also, by the Riemann-Lebesgue lemma, there exists a constant $\zeta > 0$ such that $|h(x)| < \frac{1}{2}h(0)$ if $||x||_{\infty} > \zeta$. For $\varepsilon, w > 0$, we define a function g on \mathbb{R}^d by

$$g(x) = \frac{2\varepsilon}{h(0)} h\left(\frac{x\zeta}{w}\right).$$
(19)

Here, we choose ε and w later. For $a \in \mathbb{R}^d$, we define $g_a(x)$ by g(x - a). By abuse of notation, we denote the restriction of g or g_a to \mathcal{X} by the same symbol.

The following lemma is due to (Scarlett et al., 2017):

Lemma C.6. There exist constants $C = C(d, k, \kappa)$, $C' = C'(d, k, \kappa)$, $C'' = C''(d, k, \kappa)$, $\varepsilon_0 = \varepsilon_0(d, k, \kappa) > 0$ such that the following statements are satisfied:

(i) If k is a Matérn kernel, $w = C\varepsilon^{1/\nu}$, and $0 < \varepsilon < \varepsilon_0$, then $g \in \mathcal{H}_k(\mathcal{X})$ and $\|g\|_k \leq 1$.

(ii) If k is a SE kernel, $w = C'(\log(1/\varepsilon) + C'')^{-1/2}$, and $0 < \varepsilon < \varepsilon_0$, then $g \in \mathcal{H}_k(\mathcal{X})$ and $\|g\|_k \le 1$.

Moreover, the same statements hold for translations of g_a for any $a \in \mathbb{R}^d$.

We partition the *d*-dimensional cube \mathcal{X} into M cubes R_1, \ldots, R_M with width w, where $M = \left\lfloor \left(\frac{1}{w}\right)^d \right\rfloor$. For $0 < \varepsilon < 1$ and $m = 1, \ldots, M$, we define f_m as a translation g_{a_m} of g, where a_m is a point in R_m . We define a set of environments $\{\rho_m\}_{m=1}^M$ by $\rho_m(x) = p_m(x)\delta_0 + q_m(x)\delta_1$, where δ_0 and δ_1 are Dirac delta and $p_m(x) = \alpha(1 - f_m(x)), q_m(x) = 1 - p_m(x)$. We define $\rho_0(x) = p_0(x)\delta_0 + q_0(x)\delta_1$, where $p_0(x) = \alpha(1 - f_0(x)), q_0(x) = 1 - p_0(x)$, and $f_0(x) = 0$ (constant functions). Then, by Lemma C.5, we have

$$\operatorname{CVaR}_{\alpha}(\boldsymbol{\rho}_m(x)) = \max(f_m(x), 0).$$

Lemma C.7. Assume that k is a Matérn kernel, $f_m \in \mathcal{H}_k(\mathcal{X})$, and $||f_m||_k \leq 1$ for all m. Then, the following statements hold:

- (i) $p_m, q_m \in \mathcal{H}_k(\mathcal{X})$ for all m.
- (ii) For each m, we define a linear operator $\Theta_m : \mathcal{H}_k(\mathcal{X}) \to \mathcal{H}_l(\mathcal{Y})$ by $\Theta_m(f) = \langle f, p_m \rangle_k \phi_l(0) + \langle f, q_m \rangle \phi_l(1)$. Then, there exists a constant $B = B(d, k, l, \alpha) > 0$ such that $\|\Theta_m\|_{op} \leq B$ for all m. Moreover, $\Theta_m \circ \phi_k = \mu_l \circ \rho_m$ for all m.

Proof. Since k is a Matérn kernel, constant functions belong to $\mathcal{H}_k(\mathcal{X})$ (c.f. Wendland (2004, Corollary 10.48, Theorem 10.46))⁴. Thus, by definition of p_m and q_m we have assertion (i). By $\|\Theta_m\|_{op}^2 = \sup_{\|f\|_k=1} \|\Theta_m(f)\|^2 \le \|p_m\|_k^2 + \|q_m\|_k^2 + 2\|p_m\|_k \|q_m\|_k l(0,1)$, and definitions of p_m and q_m , $\|\Theta_m\|_{op}$ is bounded by a constant $B(d,k,l,\alpha)$. Commutativity $\Theta_m \circ \phi_k = \mu_l \circ \rho_m$ is obvious by definition.

We fix a bandit algorithm $\pi = (\pi_t(x_t|y_1, \dots, y_{t-1}))_{t=1,\dots,T}$ for the kernelized CVaR bandit problem, where $\pi_t(x_t|y_1, \dots, y_{t-1})$ is a probability kernel. Let $x_{1:T} := (x_1, \dots, x_T)$ and $y_{1:T} := (y_1, \dots, y_T)$ be a random vector of selected arms and observations respectively, if the algorithm is run with an environment. We introduce the following notation as in (Scarlett et al., 2017).

- For m = 0, 1, ..., M, let \mathbf{P}_m the density function of $(x_{1:T}, y_{1:T})$ when the algorithm is run with environment ρ_m . Similarly, let $\mathbf{E}_m [\cdot]$ be the expectation with respect to $x_{1:T}$ and $y_{1:T}$ when the algorithm is run with ρ_m . We denote by $\mathbf{P}_{m,y}$ be the marginal distribution of $y_{1:T}$ and $\mathbf{E}_{m,y} [\cdot]$ is the expectation with the distribution $\mathbf{P}_{m,y}$.
- Let $N_j(y_1, \ldots, y_T) = \mathbf{E}_{x_1, \ldots, x_T} \left[\sum_{t=1}^T \mathbb{1}(x_t \in R_j) \right]$, where the expectation is taken over $x_t \sim \pi_t(x|y_1, \ldots, y_{t-1})$ for $t = 1, \ldots, T$.
- For $1 \le j \le M, 0 \le m \le M$, let $\overline{v}_m^j := \sup_{x \in R_j} f_m(x)$.
- For $1 \le j \le M, 0 \le m \le M$, let $\widetilde{v}_m^j := \sup_{x \in R_j} \max(f_m(x), 0)$.
- For $1 \leq j \leq M, 0 \leq m \leq M$, let $\overline{D}_m^j := \sup_{x \in R_j} D(\rho_0(x) || \rho_m(x))$, where $D(\cdot || \cdot)$ is the Kullback-Leibler divergence.

Since ρ_m is a Bernoulli distribution, then we easily have the following lemma:

Lemma C.8. Let $0 \le m \le M$. Then, there exists constants $\varepsilon'_0(\alpha), C(\alpha) > 0$ such that the following inequality holds for all j, m if $\varepsilon < \varepsilon'_0(\alpha)$.

$$\overline{D}_m^j \le C(\alpha) \left(\overline{v}_m^j\right)^2$$

Proof. This can be proved by $D(\rho_0(x)||\rho_m(x)) = p_0(x) \log\left(\frac{p_0(x)}{p_m(x)}\right) + q_0(x) \log\left(\frac{q_0(x)}{q_m(x)}\right)$ and the Taylor expansion of log.

⁴In the case of squared exponential kernels, constant functions does not belong to $\mathcal{H}_k(\mathcal{X})$ (Minh, 2010). This is the main reason why we assume k is a Matérn kernel in Theorem 6.1.

The following lemma follows from (Scarlett et al., 2017, Lemma 5).

Lemma C.9. We have the following statements:

- (i) $\sum_{j=1}^{M} \widetilde{v}_m^j = O(\varepsilon)$ for all $0 \le m \le M$.
- (ii) $\sum_{m=1}^{M} \widetilde{v}_m^j = O(\varepsilon)$ for all $1 \le j \le M$.
- (iii) $\sum_{m=1}^{M} \left(\overline{v}_{m}^{j}\right)^{2} = O(\varepsilon^{2})$ for all $1 \leq j \leq M$.

Proof of Theorem 6.1. In the following, we assume ε is sufficiently small i.e., $\varepsilon < \min(\varepsilon_0(d, k, \kappa), \varepsilon'_0(\alpha))$, so that assumptions in Lemma C.6, Lemma C.9, and Lemma C.8 are satisfied.

Let $V_T := \sum_{t=1}^T \text{CVaR}(\boldsymbol{\rho}_m(x_t))$. Then by Lemma C.5 and definition of ρ_m , we have $V_T = \sum_{t=1}^T \max(f_m(x_t), 0)$. Then similarly to (Scarlett et al., 2017, (66)-(69)), we have

$$\mathbf{E}_{m}\left[V_{T}\right] \leq \sum_{t=1}^{T} \sum_{j=1}^{M} \mathbf{P}_{m}(x_{t} \in R_{j}) \widetilde{v}_{m}^{j} = \sum_{j=1}^{M} \widetilde{v}_{m}^{j} \mathbf{E}_{m,y}\left[N_{j}\right]$$
$$\leq \sum_{j=1}^{M} \overline{v}_{m}^{j} \left(\mathbf{E}_{0,y}\left[N_{j}\right] + T \sqrt{\sum_{j'=1}^{M} \mathbf{E}_{0,y}\left[N_{j'}\right] \overline{D}_{m}^{j'}}\right)$$

Here the last inequality follows from (Scarlett et al., 2017, Lemma 3). Taking the average over $m = 1, \ldots, M$, we have

$$\frac{1}{M}\sum_{m=1}^{M}\mathbf{E}_{m}[V_{T}] \leq \frac{1}{M}\sum_{m=1}^{M}\sum_{j=1}^{M}\widetilde{v}_{m}^{j}\left(\mathbf{E}_{0,y}[N_{j}] + T\sqrt{\sum_{j'=1}^{M}\mathbf{E}_{0,y}[N_{j'}]\overline{D}_{m}^{j'}}\right).$$

By Lemma C.9 (ii), the first term can be bounded as follows:

$$\frac{1}{M}\sum_{m=1}^{M}\sum_{j=1}^{M}\widetilde{v}_{m}^{j}\mathbf{E}_{0,y}\left[N_{j}\right] \leq O\left(\frac{\varepsilon}{M}\right)\mathbf{E}_{0,y}\left[\sum_{j=1}^{M}N_{j}\right] = O\left(\frac{T\varepsilon}{M}\right).$$
(20)

Next, let us consider the second term. It can be bounded by the same argument to (Scarlett et al., 2017, (55)-(60)) as follows:

$$\frac{T}{M} \sum_{m=1}^{M} \sum_{j=1}^{M} \widetilde{v}_{m}^{j} \sqrt{\sum_{j'=1}^{M} \mathbf{E}_{0,y} [N_{j'}] \overline{D}_{m}^{j'}}} \leq O(T\varepsilon) \frac{1}{M} \sum_{m=1}^{M} \sqrt{\sum_{j'=1}^{M} \mathbf{E}_{0,y} [N_{j'}] \overline{D}_{m}^{j'}}} \\
\leq O(T\varepsilon) \frac{1}{M} \sum_{m=1}^{M} \sqrt{\sum_{j'=1}^{M} \mathbf{E}_{0,y} [N_{j'}] (\overline{v}_{m}^{j})^{2}}} \leq O(T\varepsilon) \sqrt{\frac{1}{M} \sum_{m=1}^{M} \sum_{j'=1}^{M} \mathbf{E}_{0,y} [N_{j'}] (\overline{v}_{m}^{j})^{2}}} \\
\leq O(T\varepsilon^{2}) \frac{1}{\sqrt{M}} \sqrt{\sum_{j'=1}^{M} \mathbf{E}_{0,y} [N_{j'}]} = O\left(\frac{\varepsilon^{2}T\sqrt{T}}{\sqrt{M}}\right).$$
(21)

By inequalities (20), (21), we have the following upper bound:

$$\frac{1}{M}\sum_{m=1}^{M} \mathbf{E}_{m}\left[V_{T}\right] = O\left(T\varepsilon\left(\frac{1}{M} + \varepsilon\sqrt{\frac{T}{M}}\right)\right).$$
(22)

Since $\text{CVaR}_{\alpha}(\rho_m(x)) = \max(f_m(x), 0)$ and f_m achieves its maximum 2ε at $x = a_m$ by construction, there exists a constant $C' = C'(d, k, \kappa, \alpha) > 0$ such that

$$\frac{1}{M}\sum_{m=1}^{M} R_{\mathrm{CVaR},\alpha}(T;\pi,\rho_m) \ge \varepsilon T \left(2 - \frac{C'}{M} - C'\varepsilon \sqrt{\frac{T}{M}}\right).$$

If we take $w = C\varepsilon^{1/\nu}$ as in Lemma C.6, then $M \to \infty$ as $\varepsilon \to 0$. Therefore, there exists a constant $\varepsilon_0'' = \varepsilon_0''(d, k, \kappa, \alpha) > 0$ such that if $\varepsilon < \varepsilon_0''$, then C'/M < 1/2. We assume ε satisfies $\varepsilon < \min(\varepsilon_0, \varepsilon_0', \varepsilon_0'')$. Then, we have

$$\frac{1}{M}\sum_{m=1}^{M} R_{\text{CVaR},\alpha}(T;\pi,\rho_m) \ge \varepsilon T\left(\frac{3}{2} - C'\varepsilon\sqrt{\frac{T}{M}}\right).$$
(23)

We take $\varepsilon = \Theta(T^{-\frac{\nu}{d+2\nu}})$. Thus, we put $\varepsilon = C''T^{-\frac{nu}{d+2\nu}}$. Then, $w = C\varepsilon^{1/\nu} = C(C'')^{1/\nu}T^{-\frac{1}{d+2\nu}}$ and there exists a constant $C^{(3)} > 0$ such that $M \ge C^{-1/d}(C'')^{-d/\nu}C^{(3)}T^{d/(d+2\nu)}$. Thus, we have

$$C'\varepsilon\sqrt{\frac{T}{M}} \le C'C^{d/2}(C'')^{1+\frac{d}{2\nu}}(C^{(3)})^{-1/2}.$$
(24)

We take C'' > 0 so that the RHS of inequality (24) is less than 1/2. Then, by inequality (23), we have

$$\frac{1}{M}\sum_{m=1}^{M} R_{\mathrm{CVaR},\alpha}(T;\pi,\rho_m) \ge \varepsilon T = \Omega(T^{\frac{d+\nu}{d+2\nu}}).$$

The assumption that ε is sufficiently small is satisfied if we take T sufficiently large. Therefore, we obtain our assertion.

Remark C.10. By the proof, we note that the statement of Theorem 6.1 remains true if we restrict the arm set to the set consisting $\{a_m\}_{m=1}^M$, where $M = \Theta(T^{d/(d+2\nu)})$.

D. Additional Results on MV

In the main article, we mainly focused on optimization of CVaR, i.e., we only provided a UCB type algorithm for MV. In this section, we consider a phased algorithm for MV and provide a lower bound for cumulative MV regret. We also conduct experiments using synthetic environments for cumulative MV regret minimization.

D.1. Phased Algorithm for MV optimization

In this section, we provide a phased algorithm for MV and a regret upper bound of it.

We outline pseudocode of the phased algorithm for MV in Algorithm 3. Here, for each phase $1 \le j \le J$, $\widehat{\mathrm{MV}}_{s}^{(j)}$ denotes the estimation of $\mathrm{MV}(\rho(x))$ defined in Proposition 5.1 but using only the observation history $\{(x_{t_j}, y_{t_j}), \ldots, (x_{s_j}, y_{s_j})\}$ in the *j*th phase, where $s_j = \min(t_{j+1} - 1, T)$.

Algorithm 3 Phased Algorithm for Kernelized MV Bandits with Finitely Many Arms

Set
$$\mathcal{X}_{1} := \mathcal{X}, J := \lfloor \log_{2}(T) \rfloor, t_{j} = 2^{j-1}$$
 for $1 \le j \le J$.
Set $\widetilde{\beta}_{k,T}^{(MV)}(\delta) := B\sqrt{\lambda} + 2\sqrt{2\log(4|\mathcal{X}|J/\delta)}$.
Set $\widetilde{\beta}_{k,T}^{(MV)}(\delta) := \mathfrak{C}\widetilde{\beta}_{k,T}^{(MV)}(\delta) + \mathfrak{C}'\left(\widetilde{\beta}_{k,T}^{(MV)}(\delta)\right)^{2}$.
for $j = 1, \ldots, J$ do
for $t = t_{j} \ldots, \min(t_{j+1} - 1, T)$ do
Play $x_{t} := \operatorname{argmax}_{x \in \mathcal{X}_{j}} \sigma_{t-1}^{(j)}(x)$ and observe y_{t} .
end for
 $s := \min(t_{j+1} - 1, T)$
 $b := \max_{x \in A_{j}} \left(\widehat{MV}_{s}^{(j)}(x) - \widetilde{\beta}_{k,T}^{(MV)}(\delta)\sigma_{s}^{(j)}(x)\right)$.
 $\mathcal{X}_{j+1} := \{x \in \mathcal{X}_{j} : \widehat{MV}_{s}^{(j)}(x) + \widetilde{\beta}_{k,T}^{(MV)}(\delta)\sigma_{s}^{(j)}(x) \ge b\}$.
end for

(**m**) | .

First, we provide a concentration inequality for $MV(\rho(x))$ similarly to Proposition 7.1.

Proposition D.1. Assume $\lambda \geq 1$. We define $\widetilde{\beta}_{k,T}^{(MV)}(\delta) = B\sqrt{\lambda} + 2\sqrt{2\log(4|\mathcal{X}|\log_2(T)/\delta)}$ and $\widetilde{\beta}_{k,T}^{(MV)}(\delta) = \mathfrak{C}\widetilde{\beta}_{k,T}^{(MV)}(\delta) + \mathfrak{C}'\left(\widetilde{\beta}_{k,T}^{(MV)}(\delta)\right)^2$. The following inequality uniformly holds for $1 \leq j \leq J$ and $x \in \mathcal{X}$ with probability at least $1 - \delta$:

$$\left| \mathrm{MV}(\boldsymbol{\rho}(x)) - \widehat{\mathrm{MV}}_{s_j}^{(j)}(x) \right| \leq \widetilde{\beta}_{k,T}^{(\mathrm{MV})}(\delta) \sigma_{s_j}^{(j)}(x).$$

Here $s_j = \min(t_{j+1} - 1, T)$.

Proof. We use the same notation as in the proof of Proposition 7.1. By (16), the following inequality holds with probability at least $1 - \delta$ for each $f \in \mathcal{H}_l(\mathcal{Y})$ with $||f||_l \leq U$, $1 \leq j \leq J$, and $x \in \mathcal{X}$:

$$\left| \mathbf{E}_{y \sim \boldsymbol{\rho}(x)} \left[f(y) \right] - \langle f, \Theta^{(j)}(x) \rangle_l \right| \le U \left(B \sqrt{\lambda} + 2\sqrt{2 \log(2/\delta)} \right) \sigma_{s_j}^{(j)}(x).$$

By applying this inequality to the case of $f = \chi_1$ and $f = \chi_2$, and taking a union bound for $f = \chi_1, \chi_2, 1 \le j \le J$, and $x \in \mathcal{X}$, we have

$$\left| m_n(\boldsymbol{\rho}(x)) - \langle \chi_n, \Theta^{(j)}(x) \rangle_l \right| \le \|\chi_n\|_l \widetilde{\beta}_{k,T}^{(\mathrm{MV})}(\delta) \sigma_{s_j}^{(j)}(x)$$

uniformly on $n \in \{1, 2\}, 1 \le j \le J$, and $x \in \mathcal{X}$. By the proof of Proposition 5.1, we have

$$\left| \operatorname{MV}\left(\boldsymbol{\rho}(x)\right) - \widehat{\operatorname{MV}}_{s_{j}}^{(j)}(x) \right| \leq \mathfrak{C}\widetilde{\beta}_{k,T}^{(\mathrm{MV})}(\delta)\sigma_{s_{j}}^{(j)}(x) + \mathfrak{C}'\left(\widetilde{\beta}_{k,T}^{(\mathrm{MV})}(\delta)\right)^{2} \left(\sigma_{s_{j}}^{(j)}(x)\right)^{2}.$$
(25)

Since $\lambda \ge 1$ and $\sup_{x \in \mathcal{X}} k(x, x) \le 1$, we have $\sigma_{s_j}^{(j)}(x) \in [0, 1]$. Thus, we have our assertion by $\left(\sigma_{s_j}^{(j)}(x)\right)^2 \le \sigma_{s_j}^{(j)}(x)$. \Box

Remark D.2. The factor $\tilde{\beta}_{k,T}^{(MV)}(\delta)$ does not depend on the information gain $\gamma_{k,T}$. In fact, we have $\tilde{\beta}_{k,T}^{(MV)}(\delta) = O\left(\sqrt{\log(\log(T)/\delta)}\right)$. Therefore, if we ignore constants and poly-logarithmic factor of T, the first term $\mathfrak{C}\tilde{\beta}_{k,T}^{(MV)}(\delta)\sigma_{s_j}^{(j)}(x)$ dominates in the right-hand side of (25). Thus unlike Proposition 5.1, Proposition D.1 does not involve the term $\left(\sigma_{s_j}^{(j)}(x)\right)^2$.

Using Proposition D.1, the following theorem can be proved in the same manner to Theorem 7.2. **Theorem D.3.** With probability at least $1 - \delta$, Algorithm 3 achieves the following regret upper bound:

$$R_{\rm MV}(T) \le 2 \sup_{x \in \mathcal{X}} {\rm MV}(\boldsymbol{\rho}(x)) + \frac{8}{\log 2} \tilde{\beta}_{k,T}^{({\rm MV})}(\delta) \sqrt{T \gamma_{k,T} \log T}.$$

D.2. Lower Bound for Cumulative MV Regret

In this subsection, we assume k is a Matérn- ν kernel. In Theorem 6.1, we provided lower bounds for cumulative CVaR regret. The following theorem provides algorithm-independent lower bounds for cumulative MV regret:

Theorem D.4. Assume c > 0, \mathcal{X} is the d-dimensional cube $[0, 1]^d$, and k is Mattern- ν kernel. Then there exist positive constants $B = B(l, c) T_0 = T_0(c, k, d)$, C = C(c, k, d) satisfying the following statement. For any $T > T_0$, there exist environments $(\rho_1, \Theta_1), \ldots, (\rho_M, \Theta_M)$ with $\|\Theta_m\|_{op} \leq B$ (for $1 \leq m \leq M$), where M = M(c, k, d, T), such that for any algorithm π , we have $\mathbf{E}[R_{MV}(T;\pi)] > CT^{\frac{\nu+d}{2\nu+d}}$. Here $\mathbf{E}[R_{MV}(T;\pi)]$ denotes the average expected cumulative regret $\frac{1}{M} \sum_{m=1}^{M} \mathbf{E}[R_{MV}(T;\pi, \rho_m, \Theta_m)]$.

Let $\{f_m\}_{1 \le m \le M}$ be the set of functions defined in §C.4. We define a set of distributions $\{\rho_m(x)\}_{1 \le m \le M}$ as follows. We choose $(p_0, q_0) \in \mathbb{R}^2$ so that $p_0, q_0 \ge 0$, $p_0 + q_0 = 1$, and $1 - c(-p_0 + q_0) > 0$ (e.g., $p_0 = \frac{c}{2c+2}, q_0 = \frac{c+2}{2c+2}$). For $1 \le m \le M$, we define $p_m(x) = p_0 + f_m(x)$ and $q_m(x) = q_0 - f_m(x)$. We take ε in (19) sufficiently small so that $p_m(x), q_m(x) \in [0, 1]$ for all $x \in \mathcal{X}$ and $1 \le m \le M$ (this is possible since $\sup_{x \in \mathbb{R}^d} |g(x)| \le 2\varepsilon$). We let $f_0(x) = 1$ (constant function). We define distribution $\rho_m(x)$ by $\rho_m(x) = p_m(x)\delta_1 + q_m(x)\delta_0$. Since $\rho_0(x) = p_0\delta_1 + q_0\delta_0$ does not depend on x, we denote $\rho_0(x)$ by ρ_0 . Then, by definition, we obtain the following:

$$MV(\boldsymbol{\rho}_m(x)) = MV(\rho_0) + (1 - c(-p_0 + q_0)) f_m(x) + cf_m^2(x).$$
(26)

We can prove the following lemma similarly to C.7.

Lemma D.5. Assume that k is a Matérn kernel, $f_m \in \mathcal{H}_k(\mathcal{X})$, and $||f_m||_k \leq 1$ for all m^5 Then, the following statements hold:

- (i) $p_m, q_m \in \mathcal{H}_k(\mathcal{X})$ for all m.
- (ii) For each m, we define a linear operator $\Theta_m : \mathcal{H}_k(\mathcal{X}) \to \mathcal{H}_l(\mathcal{Y})$ by $\Theta_m(f) = \langle f, p_m \rangle_k \phi_l(0) + \langle f, q_m \rangle \phi_l(1)$. Then, there exists a constant B = B(d, k, l, c) > 0 such that $\|\Theta_m\|_{op} \leq B$ for all m. Moreover, $\Theta_m \circ \phi_k = \mu_l \circ \rho_m$ for all m.

We define notations $x_1, \ldots, x_t, y_1, \ldots, y_T, N_j(y_1, \ldots, y_T), \overline{v}_m^j$, and \overline{D}_m^j in the same way as in §C.4. We define \widetilde{v}_m^j as follows:

$$\widetilde{v}_m^j := \sup_{x \in R_j} MV(\boldsymbol{\rho}_m(x)) - MV(\rho_0) = \sup_{x \in R_j} \left\{ (1 - c(-p_0 + q_0)) f_m(x) + c f_m^2(x) \right\}.$$

The following lemma can be proved in the same way as Lemma C.8.

Lemma D.6. There exists a constant $\varepsilon' = \varepsilon'(c) > 0$ such that the following statement holds. If $\varepsilon < \varepsilon'$, then we have $\overline{D}_m^j = O\left(\left(\overline{v}_m^j\right)^2\right)$ for all $0 \le m \le M$ and $1 \le j \le M$.

Since $\sup_{x \in \mathcal{X}} |f_m(x)| = 2\varepsilon$ and we take ε sufficiently small (e.g., $2\varepsilon < 1$), we have $f_m(x)^2 = O(|f_m(x)|)$ uniformly for $x \in \mathcal{X}$. Thus, the same statements to Lemma C.9 hold:

Lemma D.7. (i) $\sum_{j=1}^{M} \widetilde{v}_m^j = O(\varepsilon)$ for all $0 \le m \le M$.

- (ii) $\sum_{m=1}^{M} \tilde{v}_m^j = O(\varepsilon)$ for all $1 \le j \le M$.
- (iii) $\sum_{m=1}^{M} \left(\overline{v}_{m}^{j}\right)^{2} = O(\varepsilon^{2})$ for all $1 \le j \le M$.

Remark D.8. Although $f_m(x)$ can take negative values, the proof of (Scarlett et al., 2017, Lemma 5) works even if we replace $f_m(x)$ by $|f_m(x)|$ in the definition of \overline{v}_m^j . Thus, by $\overline{v}_m^j = O(\sup_{x \in R_j} |f_m(x)|)$ and $\widetilde{v}_m^j = O(\sup_{x \in R_j} |f_m(x)|)$, we have the statements of the lemma by (Scarlett et al., 2017, Lemma 5).

Proof of Theorem D.4. We take ε sufficiently small so that the assumptions of Lemma C.6, Lemma D.5, and Lemma D.7 are satisfied.

We put $V_T = \sum_{t=1}^T MV(\boldsymbol{\rho}(x_t))$. Then, it follows that

$$\mathbf{E}_m[V_T] \le T \operatorname{MV}(\rho_0) + \sum_{t=1}^T \sum_{j=1}^M \mathbf{P}_m(x_t \in R_j) \widetilde{v}_m^j.$$

Then, by the same argument as (22), we have

$$\frac{1}{M}\sum_{m=1}^{M} \mathbf{E}_{m} \left[V_{T} \right] = T \operatorname{MV}(\rho_{0}) + O\left(T\varepsilon \left(\frac{1}{M} + \varepsilon \sqrt{\frac{T}{M}} \right) \right)$$

By (26) and $\sup_{x \in \mathcal{X}} f_m(x) = 2\varepsilon$, we have $\sup_{x \in \mathcal{X}} MV(\rho(x)) = MV(\rho_0) + 2(1 - c(-p_0 + q_0))\varepsilon + 4\varepsilon^2$. Therefore, we have

$$\frac{1}{M} \sum_{m=1}^{M} R_{\rm MV}(T; \pi, \rho_m) \ge \varepsilon T \left\{ 2(1 - (-p_0 + q_0)) + 4\varepsilon - \frac{C'}{M} - C'\varepsilon \sqrt{\frac{T}{M}} \right\}$$
$$\ge \varepsilon T \left\{ 2(1 - (-p_0 + q_0)) - \frac{C'}{M} - C'\varepsilon \sqrt{\frac{T}{M}} \right\}.$$

⁵This condition can be satisfied for appropriate choices of ε and w by Lemma C.6.

Then, by the same argument as in the proof of Theorem 6.1, we obtain

$$\frac{1}{M} \sum_{m=1}^{M} R_{\rm MV}(T;\pi,\rho_m) = \Omega(T^{\frac{d+\nu}{d+2\nu}}).$$

D.3. Experiments for Cumulative MV Regret Minimization

In this section, we empirically compare MVPKE-UCB to RAHBO (Makarova et al., 2021), which is also a kernelized bandit algorithm for MV optimization. We use the same synthetic environments introduced in §8 and conducted experiments for c = 1, where c is the parameter of MV. To evaluate RAHBO in our environments, we used our own implementation of it. RAHBO (Makarova et al., 2021) repeatedly observe outputs at the same arm. In our experiments, we took the number of repetition (denoted by k in (Makarova et al., 2021)) as 5 as in the Random Forest experiment in (Makarova et al., 2021). As in (Makarova et al., 2021), we assume parameters β_t , β_t^{var} of RAHBO are constants. Similarly, we assume $\mathfrak{C}\beta_{k,t}^{(MV)}(\delta)$ and $\mathfrak{C}'\left(\beta_{k,t}^{(MV)}(\delta)\right)^2$ are constants in the experiments and denote them by β_1 and β_2 respectively. We search optimal values of β_t , β_t^{var} , β_1 , β_2 with in geometric sequence $\{\beta^{(1)}, \ldots, \beta^{(4)}\}$ with $\beta^{(1)} = 10^{-2}$ and $\beta^{(4)} = 1.0$ in different synthetic environments from those for evaluation using the first 200 rounds and estimations of MV. To simplify the parameter search, we assume $\overline{\rho} = 1.0$ for RAHBO. We set $\lambda = 1$ for both the algorithms.

We compute normalized cumulative MV regret for both environments and the results are displayed in Figure 3. Here we normalize cumulative MV regret so that uniform random policy incurs cumulative MV regret 1 at T = 1000. In this experimental settings, for both kinds of environments, we found that RAHBO incurs larger cumulative regret than MVPKE-UCB in average. Especially, the experimental results indicate that RAHBO is not better than the uniform random policy for LogNormal environments. We suspect that this is because LogNormal distributions with mean 0 do not satisfy sub-Gaussian assumption, but Makarova et al. (2021) assumes sub-Gaussian noise for their theoretical analysis.



Figure 3. Cumulative MV Regret in Synthetic Environments

E. Details of the Experiments

This section details our experimental setting. We used Intel(R) Core(TM) i9-9920X CPU for our experiments. For each m = 1, ..., 10, we constructed functions μ_m and σ_m as follows. For each m, we randomly select n = 100 points $\{\xi_1, ..., \xi_n\}$ from \mathcal{X} and real numbers $a_1, ..., a_n$ from the interval (-1, 1), define $\mu_m(x)$ by a constant multiple of $\sum_{i=1}^n a_i k(x, x_i)$ so that $\|\mu_m\|_k = 1$. We similarly and independently define a random function $\tilde{\sigma}_m(x)$ and define $\sigma_m(x)$ as $\sqrt{\sigma_0 + \tilde{\sigma}_m^2(x)}$, where $\sigma_0 = 10^{-3}$.

Our baseline algorithm is IGP-UCB (Chowdhury & Gopalan, 2017), and it selects arm x with the largest UCB $\mu_{k,t}(x) + \beta_{\text{IGP},k,t}(\delta)\sigma_{k,t}(x)$ at t + 1th round, where $\mu_{k,t}(x) = (y_1, \ldots, y_t) (\mathbf{k}(x_{1:t}, x_{1:t}) + \lambda \mathbf{1}_t)^{-1} \mathbf{k}(x_{1:t}, x)$ and $\beta_{\text{IGP},k,t}(\delta) = B + R\sqrt{2(\gamma_{k,t} + 1 + \log(1/\delta))}$ with parameters B, R > 0. To ease computation and parameter tuning, we made the following modification of $\beta_{\text{IGP},k,t}(\delta)$ and $\beta_{k,t}^{(\text{CV})}(\delta)$. In the definitions of $\beta_{\text{IGP},k,t}(\delta)$ and $\beta_{k,t}^{(\text{CV})}(\delta)$, we replace the

maximum information gain $\gamma_{k,t}$ by $\tilde{\gamma}_{k,t} := \frac{1}{2} \log \det \left(1_t + \lambda^{-1} \mathbf{k}(x_{1:t}, x_{1:t}) \right)$. The value $\tilde{\gamma}_{k,t}$ can be updated in $O(t^2)$ time by Lemma C.1. In addition, it is not difficult to see that IGP-UCB and CVPKE-UCB have the same regret upper bound if we replace $\gamma_{k,t}$ by $\tilde{\gamma}_{k,t}$ (c.f., Chowdhury & Gopalan (2017)). Since the term $R\sqrt{2(\gamma_{k,t} + 1 + \log(1/\delta))}$ is dominant in the definition of $\beta_{\text{IGP},k,t}(\delta)$, we let B = 0.0 for the experiments. Similarly, we let B = 0.0 in the definition of $\beta_{k,t}^{(\text{MV})}$. To tune parameter R for IGP-UCB, for both two kinds of distributions, we generated five environments which are different from the 10 environments for evaluation, and used the first 200 rounds for tuning. We tuned R for IGP-UCB so that IGP-UCB maximizes $\sum_{t=1}^{200} y_t$ since it does not have an estimation of CVaR. We searched R in geometric sequence $\{R_1, \ldots, R_{10}\}$ with $R_1 = 10^{-2}$ and $R_{10} = 10^{-1}$. We tuned R for CVPKE-UCB and tuned R in geometric sequence $\{R'_1, \ldots, R'_{10}\}$ with $R'_1 = 10^{-3}$ and $R_{10} = 10^{-1}$.

Since we generated $\mu_m(x)$ and $\sigma_m(x)$ using random elements of $\mathcal{H}_k(\mathcal{X})$, we assumed the length scale of the kernel k was known. We note that for any $f \in \mathcal{H}_l(\mathcal{Y})$, the function $x \mapsto \mathbf{E}_{y \sim \rho(x)}[f(y)] = \langle \Theta \circ \phi_k(x), f \rangle_l = \langle \phi_k(x), \Theta^*(f) \rangle_k$ belongs to $\mathcal{H}_k(\mathcal{X})$. Therefore, one possible way to tune the length scale is to apply the existing method (Rasmussen, 2003, chapter 5) to transformed observation history $(x_1, f(y_1)), \ldots, (x_t, f(y_t))$ with $f \in \mathcal{H}_l(\mathcal{Y})$. If we take f(y) = y for $y \in \mathcal{Y}$, then the method is the same as the tuning method in the conventional setting.