High-Dimensional Experimental Design and Kernel Bandits

Romain Camilleri [1] **Julian Katz-Samuels** [2] **Kevin Jamieson** [1]

Abstract

In recent years methods from optimal linear experimental design have been leveraged to obtain state of the art results for linear bandits. A design returned from an objective such as G-optimal design is actually a probability distribution over a pool of potential measurement vectors. Consequently, one nuisance of the approach is the task of converting this continuous probability distribution into a discrete assignment of N measurements. While sophisticated rounding techniques have been proposed, in d dimensions they require N to be at least d, $d\log(\log(d))$, or d^2 based on the sub-optimality of the solution. In this paper we are interested in settings where N may be much less than d, such as in experimental design in an RKHS where d may be effectively infinite. In this work, we propose a rounding procedure that frees N of any dependence on the dimension d, while achieving nearly the same performance guarantees of existing rounding procedures. We evaluate the procedure against a baseline that projects the problem to a lower dimensional space and performs rounding which requires N to just be at least a notion of the effective dimension. We also leverage our new approach in a new algorithm for kernelized bandits to obtain state of the art results for regret minimization and pure exploration. An advantage of our approach over existing UCB-like approaches is that our kernel bandit algorithms are also robust to model misspecification.

1. Introduction

This work studies a non-parametric multi-armed bandit game through the lens of experimental design. Fix a finite set of measurements $\mathcal{X} \subset \mathbb{R}^d$ and a function $\mu : \mathcal{X} \to \mathbb{R}$. We consider the following game between a learner and nature: at each time $t = 1 \dots T$, the learner requests $x_t \in \mathcal{X}$ and nature immediately reveals

$$y_t = \mu_{x_t} + \xi_t$$

where $\{\xi_t\}_{t=1}^T$ is a sequence of independent, mean-zero random variables with bounded variance. We are interested in two objectives:

Regret minimization In this setting, we evaluate the performance of an algorithm choosing actions $\{x_t\}_{t=1}^T$ by its cumulative regret: $R_T = \max_{x\in\mathcal{X}} \sum_{t=1}^T (\mu_x - \mu_{x_t})$.

Pure exploration in the PAC setting For a tolerance $\epsilon \geq 0$ and confidence level $\delta \in (0, 1)$, the aim of the learner in pure exploration is to sequentially take samples until a learner-defined stopping criterion is met, at which time the learner outputs an arm $\widehat{x} \in \mathcal{X}$ such that $\mu_{\widehat{x}} \geq \max_{x\in\mathcal{X}} \mu_x - \epsilon$ with probability at least $1 - \delta$.

To aid us in our objectives, we assume some structure on the reward function μ.

Assumption 1. *There exists a known feature map $\phi : \mathbb{R}^d \mapsto \mathcal{H}$ that maps each $x \in \mathcal{X}$ to a (possibly infinite dimensional) Hilbert space $\mathcal{H}$, and moreover, there exists a $\theta_* \in \mathcal{H}$ and $h \geq 0$ such that $\max_{x\in\mathcal{X}} |\mu_x - \langle\theta_*, \phi(x)\rangle_{\mathcal{H}}| \leq h$.*

Consequently, if h is not too big, the expected value of each of the observations y_t is nearly a linear function of its associated features $\phi(x_t)$. We say the model is *misspecified* when $h > 0$, and otherwise the setting is well-specified and reduces to the classical stochastic setting when $h = 0$.

Assumption 2. *Rewards are bounded $\max_{x\in\mathcal{X}} |\mu_x| \leq B$.*

Assumption 3. *For every time t, the additive stochastic noise ξ_t is independent, mean-zero with $\mathbb{E}[\xi_t^2] \leq \sigma^2$.*

While we assume the learner knows B and σ^2, we assume that the learner *does not* know the extent of the model misspecification $h \geq 0$. Note that we do *not* assume ξ_t is bounded, indeed, it can even be heavy tailed.

1.1. Elimination algorithms and experimental design

Whether the model is misspecified ($h > 0$) or not ($h = 0$), a popular class of algorithms for both the objectives of

[1]Paul G. Allen School of Computer Science and Engineering, University of Washington, Seattle, WA [2]University of Wisconsin, Madison, WI. Correspondence to: Romain Camilleri <camilr@cs.washington.edu>.

Proceedings of the 38^{th} International Conference on Machine Learning, PMLR 139, 2021.

regret minimization and pure exploration is known as *elimination algorithms*. Elimination algorithms proceed in stages, maintaining a set $\widehat{\mathcal{X}} \subset \mathcal{X}$ of candidates that may achieve $\max_{x\in\mathcal{X}} \mu_x$ given all previous observations. At the beginning of the stage $\ell \geq 1$ the algorithm decides which measurements to take, nature reveals the observations, and the stage ends by constructing an estimate $\widehat{\mu}_{(\cdot)}$ of $\mu_{(\cdot)}$ and removing all elements $x \in \widehat{\mathcal{X}}$ from $\widehat{\mathcal{X}}$ where $\max_{x'\in\widehat{\mathcal{X}}} \widehat{\mu}_{x'} - \widehat{\mu}_x > \epsilon_\ell$. This process is repeated indefinitely in the case of regret minimization, or until $\widehat{\mathcal{X}}$ contains a single element in the case of pure exploration. To be as effective as possible at discarding as many candidates as possible in the elimination stage (without discarding the best arm), a natural strategy of selecting how many and which measurements to take in the beginning of the round is to select $x_1, \ldots, x_n \in \mathcal{X}$ to accurately estimate the differences of the estimates

$$\max_{x,x'\in\widehat{\mathcal{X}}} (\widehat{\mu}_{x'} - \widehat{\mu}_x) - (\mu_{x'} - \mu_x) \leq \epsilon_\ell. \tag{1}$$

If $x_* := \arg\max_{x\in\mathcal{X}} \mu_x$ and $x_* \in \widehat{\mathcal{X}}$ at the start of the round, then we have that x_* will not be eliminated at the end since

$$\max_{x'\in\widehat{\mathcal{X}}} \widehat{\mu}_{x'} - \widehat{\mu}_{x_*} \leq \max_{x'\in\widehat{\mathcal{X}}} \mu_{x'} - \mu_{x_*} + \epsilon_\ell \leq \epsilon_\ell.$$

And moreover, it is straightforward to show that after the discarding step of stage ℓ, $\max_{x\in\widehat{\mathcal{X}}} \mu_{x_*} - \mu_x \leq 2\epsilon_\ell$. To guide our choice of $x_1, \ldots, x_n \in \mathcal{X}$ to achieve (1), we exploit the assumed (nearly) linear model of above.

1.2. Optimal experimental design and the problem of rounding continuous designs

This section introduces the method of experimental design with the goal of achieving (1) by taking as few total samples as possible. Shortly, we will consider the case when $h > 0$ and ϕ is an arbitrary feature map. But for now, let us make the simplifying assumption that $h = 0$, ϕ is the identity map so that $\mu_x = \langle \theta_*, x\rangle$, and $\xi_t \sim \mathcal{N}(0, \sigma^2)$. Thus, if at time t we select $x_t \in \mathcal{X} \subset \mathbb{R}^d$ we observe $\langle \theta_*, x_t\rangle + \xi_t$. Suppose we observed pairs $\{(x_t, y_t)\}_{t=1}^T$ where each $x_t \in \mathcal{X}$ was chosen independently of any y_s for $s \leq t$. If we wished to achieve (1) for $\widehat{\mathcal{X}} \subset \mathcal{X}$ with $\mu_x = \langle x, \theta_*\rangle$, perhaps the most natural way forward would be to compute the least squares estimator $\widehat{\theta}_{LS} = \arg\min_\theta \sum_{t=1}^T (y_t - \langle x_t, \theta\rangle)^2$, and set $\widehat{\mu}_x = \langle \widehat{\theta}_{LS}, x\rangle$. Then (1) is equivalent to $\max_{v\in\mathcal{V}} \langle \widehat{\theta}_{LS} - \theta_*, v\rangle \leq \epsilon_\ell$ with $\mathcal{V} = \widehat{\mathcal{X}} - \widehat{\mathcal{X}}$. By a standard sub-Gaussian tail-bound (Lattimore & Szepesvári, 2020), we have with probability at least $1-\delta$ that for all $v \in \mathcal{V} \subset \mathbb{R}^d$

$$|\langle v, \widehat{\theta}_{LS} - \theta_*\rangle| \leq \|v\|_{(\sum_{t=1}^T x_t x_t^\top)^{-1}} \sqrt{2\sigma^2 \log(2|\mathcal{V}|/\delta)}, \tag{2}$$

where we adopt the notation $\|z\|_A = \sqrt{z^\top A z}$ for any $z \in \mathbb{R}^d$ and symmetric semi-definite positive A. Note that this error bound only depends on those x_t measurements that we choose *before* any responses y_t are observed. This allows us to plan, that is, choose the T measurement vectors to minimize the RHS of (2). Unfortunately, this minimization problem is known to be NP-hard (Pukelsheim, 2006; Allen-Zhu et al., 2017). As a consequence, approximation algorithms based on the relaxation

$$\bar{\lambda} = \text{argmin}_{\lambda\in\triangle_\mathcal{X}} \max_{v\in\mathcal{V}} v^\top \left(\sum_{x\in\mathcal{X}} \lambda_x x x^\top\right)^{-1} v \tag{3}$$

have been proposed. These first solve for $\bar{\lambda}$ and "round" this to a discrete allocation of measurements.

Deterministic rounding Perhaps the simplest scheme is to obtain a solution $\bar{\lambda}$ of (3) and then sample $x \in \mathcal{X}$ exactly $\lceil \bar{\lambda}_x T \rceil$ times. In the worst case, this will result in $|\text{support}(\bar{\lambda})|$ additional measurements than the intended T. Caratheodory's theorem provides a polynomial-time algorithm for constructing $\tilde{\lambda} \in \triangle_\mathcal{X}$ such that $\sum_{x\in\mathcal{X}} \tilde{\lambda}_x x x^\top = \sum_{x\in\mathcal{X}} \bar{\lambda}_x x x^\top$ and $|\text{support}(\tilde{\lambda})| \leq (d+1)d/2$. However, more sophisticated rounding procedures exist. (Allen-Zhu et al., 2017) inflates the RHS of (2) by a constant factor while only requiring that $T = \Omega(d)$. When $\mathcal{V} = \mathcal{X}$, another strategy is to solve the optimization problem (3) with a Frank-Wolfe style algorithm that is terminated only after $O(d \log\log(d))$ iterations so that the rounding according to the naive ceiling operation only inflates T by the number of iterations which is $O(d \log\log(d))$ (Todd, 2016).

Stochastic rounding Another basic rounding algorithm simply samples $x_1, \ldots, x_T \sim \bar{\lambda}$. Unfortunately, using the least squares estimator $\widehat{\theta}_{LS}$, we may have that $\sum_{t=1}^T x_t x_t^\top$ deviates dramatically from $T\sum_{x\in\mathcal{X}} \bar{\lambda}_x x x^\top$ for moderate T, thus any guarantees require T to be poly(d) and moreover, performance relies on the spectrum of $\sum_{x\in\mathcal{X}} \bar{\lambda}_x x x^\top$ (Rizk et al., 2020). As a consequence, (Tao et al., 2018) proposed using the inverse propensity score (IPS) estimator $\widehat{\theta}_{IPS} := (\sum_{x\in\mathcal{X}} \bar{\lambda}_x x x^\top)^{-1} (\frac{1}{T}\sum_{t=1}^T x_t y_t)$. From (Tao et al., 2018), with probability at least $1-\delta$ we have for all $v \in \mathcal{V}$ simultaneously

$$|\langle v, \widehat{\theta}_{IPS} - \theta_*\rangle| \leq \sqrt{\frac{2\sigma^2 \|v\|^2_{A(\bar{\lambda})^{-1}} \log(2|\mathcal{V}|/\delta)}{T}} + \frac{\log(2|\mathcal{V}|/\delta)(1 + \max_{x\in\mathcal{X}} |v^\top A(\bar{\lambda})^{-1} x|)}{T}. \tag{4}$$

where $A(\lambda) := \sum_{x\in\mathcal{X}} \lambda_x x x^\top$. The second term of (4) accounts for potentially rare but large deviations of size $\max_{x\in\mathcal{X}} |v^\top A(\bar{\lambda})^{-1} x|$. Sadly, this second term is cumbersome in analyses since it can dominate the first term, and it cannot be removed in the worst-case. A final class of algorithms rely on *proportional volume sampling*, or sampling from a determinantal point process (DPP), but are limited to specific optimality criteria (Nikolov et al., 2019; Derezinski et al., 2020).

1.3. Main contributions

The main contributions of this paper include a novel scheme for experimental design and its application to kernel bandits.

- We propose an estimator $\widehat{\theta}_{RIPS}$ that overcomes many of the shortcomings of the prior art reviewed in Section 1.2 for $h = 0$ and $\phi \equiv$ identity. For any fixed $\theta_* \in \mathbb{R}^d$, $\mathcal{V} \subset \mathbb{R}^d, \mathcal{X} \subset \mathbb{R}^d$, $\lambda \in \triangle_{\mathcal{X}}$, and $T \in \mathbb{N}$, if T samples are drawn randomly according to λ to construct $\widehat{\theta}_{RIPS}$, then with probability at least $1-\delta$ we have for all $v \in \mathcal{V}$

$$|\langle v, \widehat{\theta}_{RIPS} - \theta_* \rangle| \leq \|v\|_{(\sum_{x\in\mathcal{X}} \lambda_x x x^\top)^{-1}} \sqrt{\frac{c(\sigma^2 + B^2)\log(2|\mathcal{V}|/\delta)}{T}}$$

  for an absolute constant c. Note that our method puts no restrictions on T but matches the ideal discrete allocation of (2) up to a constant by realizing that

$$\frac{\inf_{\lambda\in\triangle_{\mathcal{X}}} \max_{v\in\mathcal{V}} \|v\|_{(\sum_{x\in\mathcal{X}} T\lambda_x x x^\top)^{-1}}}{\min_{\{x_t\}_{t=1}^T \in \mathcal{X}} \max_{v\in\mathcal{V}} \|v\|_{(\sum_{t=1}^T x_t x_t^\top)^{-1}}} \leq 1.$$

  We also note that we only assume the stochastic noise has bounded variance and do not rule out heavy-tailed distributions. The estimator $\widehat{\theta}_{RIPS}$ is a special case of our more general estimator.

- We extend our estimator to the misspecified setting where $h \geq 0$ and to use feature maps $\phi : \mathbb{R}^d \to \mathcal{H}$ for an RKHS $\mathcal{H}$. When $\mathcal{H}$ can represent a high or even an infinite dimensional space, restrictions on T based on the dimension start to become paramount. For any fixed $\theta_* \in \mathcal{H}$, $\mathcal{V} \subset \mathcal{H}, \mathcal{X} \subset \mathbb{R}^d$, $\lambda \in \triangle_{\mathcal{X}}$, $T \in \mathbb{N}$, and $\gamma \geq 0$, if T samples are drawn randomly according to λ to construct $\widehat{\theta}_{RIPS}(\gamma)$, then with probability at least $1-\delta$ we have for all $v \in \mathcal{V}$

$$|\langle v, \widehat{\theta}_{RIPS}(\gamma) - \theta_* \rangle| \leq \|v\|_{(\sum_{x\in\mathcal{X}} \lambda_x \phi(x)\phi(x)^\top + \gamma I)^{-1}} \times \left(\sqrt{\gamma}\|\theta_*\|_2 + h + \sqrt{\frac{c(\sigma^2+B^2)\log(2|\mathcal{V}|/\delta)}{T}}\right).$$

  Note that since $\mathcal{H}$ may be infinite-dimensional, the estimator $\widehat{\theta}_{RIPS}(\gamma)$ is constructed implicitly and is implemented through kernel evaluations only.

- We empirically compare $\widehat{\theta}_{RIPS}(\gamma)$ to the sampling and estimator pairs of Section 1.2 and show that $\widehat{\theta}_{RIPS}(\gamma)$ is competitive on both finite dimensional G-optimal design as well as its regularized RKHS variant sometimes called Bayesian experimental design.

- We employ $\widehat{\theta}_{RIPS}(\gamma)$ in a novel elimination style algorithm for kernel bandits. Our regret bounds match state of the art results in the well-specified setting, and are the first linear bounds that we are aware of for the misspecified setting. In addition, we state an instance-dependent pure-exploration result for identifying an ϵ-good arm with probability at least $1-\delta$ that compares favorably to known lower bounds. One advantage of our algorithm over prior kernel bandits and Bayesian Optimization algorithms (Srinivas et al., 2009; Valko et al., 2013; Frazier, 2018) is that our approach naturally allows for taking batches of pulls per round.

2. Robust Inverse Propensity Score (RIPS) estimator

In this section we introduce the $\widehat{\theta}_{RIPS}$ estimator. In finite dimensions, our estimator first constructs $\widehat{\theta}_{IPS}$ but then to avoid the large deviations term of (4) applies robust mean estimation on each $\langle v, \theta_* \rangle$ to obtain a $\widehat{\theta}_{RIPS}$ which is consistent with all of these estimates. When we move to an RKHS setting, we add regularization to avoid vacuous bounds and account for the introduced bias. The bias of misspecification is handled similarly. We begin with robust mean estimation.

Definition 1. *Let $X_1, \ldots, X_n$ be i.i.d. random variables with mean $\bar{x}$ and variance ν^2. Let $\delta \in (0,1)$. We say that $\widehat{\mu}(X_1, \ldots, X_n)$ is a* δ-robust estimator *if there exist universal constants $c_1, c_0 > 0$ such that if $n \geq c_1 \log(1/\delta)$, then with probability at least $1-\delta$*

$$|\widehat{\mu}(\{X_t\}_{t=1}^n) - \bar{x}| \leq c_0 \sqrt{\frac{\nu^2 \log(1/\delta)}{n}}.$$

Examples of δ-robust estimators include the median-of-means estimator and Catoni's estimator (Lugosi & Mendelson, 2019). This work employs the use of the Catoni estimator which satisfies $|\widehat{\mu}(\{X_t\}_{t=1}^n) - \bar{x}| \leq \sqrt{\frac{2\nu^2\log(1/\delta)}{n - 2\log(1/\delta)}}$ for $n > 2\log(1/\delta)$ which leads to an optimal leading constant as $n \to \infty$. We will use a separate robust mean estimate for each $v \in \mathcal{V}$. In particular, to estimate $\langle v, \theta_* \rangle$ we use $\widehat{\mu}(\{v^\top A^{(\gamma)}(\lambda)^{-1}\phi(x_t) y_t\}_{t=1}^T)$ where

$$A^{(\gamma)}(\lambda) := \sum_{x\in\mathcal{X}} \lambda_x \phi(x)\phi(x)^\top + \gamma I. \tag{5}$$

Our RIPS procedure for experimental design in an RKHS is presented in Figure 1. It has the following guarantee.

Theorem 1. *Fix any finite sets $\mathcal{X} \subset \mathbb{R}^d$ and $\mathcal{V} \subset \mathcal{H}$, feature map $\phi : \mathbb{R}^d \to \mathcal{H}$, number of samples τ and regularization $\gamma > 0$. If the RIPS procedure of Figure 1 is run with $\frac{\delta}{|\mathcal{V}|}$-robust mean estimator $\widehat{\mu}(\cdot)$ and if $\tau \geq c_1 \log(|\mathcal{V}|/\delta)$ then*

Algorithm 1 RIPS for Experimental Designs in an RKHS

Input: Finite sets $\mathcal{X} \subset \mathbb{R}^d$ and $\mathcal{V} \subset \mathcal{H}$, feature map $\phi : \mathbb{R}^d \to \mathcal{H}$, number of samples τ, regularization $\gamma > 0$, robust mean estimator $\widehat{\mu} : \mathbb{R}^* \to \mathbb{R}$

$$\lambda^* := \arg\min_{\lambda \in \triangle_{\mathcal{X}}} \max_{v \in \mathcal{V}} \|v\|_{\left(\sum_x \lambda_x \phi(x)\phi(x)^\top + \gamma I\right)^{-1}} \quad (6)$$

Randomly draw $\widetilde{x}_1, \ldots, \widetilde{x}_\tau$ from $\mathcal{X}$ according to λ^*
Set $W^{(v)} = \widehat{\mu}(\{v^\top A^{(\gamma)}(\lambda^*)^{-1}\phi(\widetilde{x}_t)\widetilde{y}_t\}_{t=1}^\tau)$
Set $\widehat{\theta} := \arg\min_\theta \max_{v \in \mathcal{V}} \frac{|\langle \theta, v\rangle - W^{(v)}|}{\|v\|_{(\sum_{x\in\mathcal{X}} \lambda^*_x \phi(x)\phi(x)^\top + \gamma I)^{-1}}}$

Return: $\{W^{(v)}\}_{v\in\mathcal{V}}, \widehat{\theta}$

Figure 1. In this work, we assume each element in $\mathcal{V}$ is a linear combination of $\phi \circ \mathcal{X}$ which makes all quantities well-defined and can be computed using kernel evaluations $k(x, x') := \langle \phi(x), \phi(x')\rangle_{\mathcal{H}}$. Moreover, Equation 6 is convex with gradients that can be computed using kernel evaluations (See Section 2.3).

with probability at least $1 - \delta$*, we have*

$$\max_{v\in\mathcal{V}} \frac{|W^{(v)} - \langle \theta_*, v\rangle|}{\|v\|_{(\sum_{x\in\mathcal{X}} \lambda_x \phi(x)\phi(x)^\top + \gamma I)^{-1}}} \le \sqrt{\gamma}\|\theta_*\|_2 + h + c_0\sqrt{\tfrac{(B^2+\sigma^2)}{\tau}\log(2|\mathcal{V}|/\delta)},$$

Moreover, $W^{(v)} = \widehat{\mu}(\{v^\top A^{(\gamma)}(\lambda)^{-1}\phi(x_t)y_t\}_{t=1}^\tau)$ *can be replaced by* $\langle \widehat{\theta}, v\rangle$ *by multiplying the RHS by a factor of* 2*.*

Proof sketch. Due to the regularization and potential misspecification if $h > 0$, each $v^\top A^{(\gamma)}(\lambda)^{-1}\phi(x_t)y_t$ is biased. Thus, we apply the guarantee of $W^{(v)} = \widehat{\mu}(\{v^\top A^{(\gamma)}(\lambda)^{-1}\phi(x_t)y_t\}_{t=1}^\tau)$ to the expectation of its arguments. The triangle inequality followed by repeated applications of Cauchy-Schwartz yields

$$\begin{aligned}|W^{(v)} - \langle v, \theta_*\rangle| \le& |W^{(v)} - \mathbb{E}[v^\top A^{(\gamma)}(\lambda)^{-1}\phi(x_1)y_1]| \\ &+ |\mathbb{E}[v^\top A^{(\gamma)}(\lambda)^{-1}\phi(x_1)y_1] - \langle v, \theta_*\rangle| \\ \le& c_0\sqrt{\frac{\nu^2\log(1/\delta)}{\tau}} + \sqrt{\gamma}\|\theta_*\|_2 + h\end{aligned}$$

where we obtain an upper bound on the variance ν^2 by

$$\begin{aligned}\mathbb{V}\mathrm{ar}(v^\top A^{(\gamma)}(\lambda)^{-1}\phi(x_1)y_1) \le& \mathbb{E}[(v^\top A^{(\gamma)}(\lambda)^{-1}\phi(x_1)y_1)^2] \\ =& \mathbb{E}\left[\left(v^\top A^{(\gamma)}(\lambda)^{-1}\phi(x_1)\right)^2 \mu_{x_1}^2\right] \\ +& \mathbb{E}\left[\left(v^\top A^{(\gamma)}(\lambda)^{-1}\phi(x_1)\right)^2 \xi_1^2\right] \le (B^2+\sigma^2)\|v\|^2_{A^{(\gamma)}(\lambda)^{-1}}.\end{aligned}$$

□

2.1. Practical implementation of the algorithms

The construction of $\widehat{\theta}$ in the algorithms may–at first glance–look confusing in the infinite dimensional case. In actuality, the equivalent dual representation $\widehat{\theta} = \sum_{i=1}^{|\mathcal{X}|} \alpha_i \phi(x_i)$ would be used. That is, the potentially infinite dimensional object $\widehat{\theta}$ is represented by a finite dimensional weight vector $\alpha \in \mathbb{R}^{|\mathcal{X}|}$. With that, the optimizations in the algorithms (e.g., to compute the RIPS estimator) are over the dual vector $\alpha \in \mathbb{R}^{|\mathcal{X}|}$, and inner products $\langle \widehat{\theta}, v\rangle = \sum_{i=1}^{|\mathcal{X}|} \alpha_i \langle \phi(x_i), v\rangle$ are computed using the kernel matrix of $\mathcal{X}$ since in all instances of v used in the algorithms, v is a linear combination of $\{\phi(x)\}_{x\in\mathcal{X}}$.

2.2. Comparison to IPS estimator

Note the difference between the bound of RIPS in Theorem 1 with the bound of the IPS estimator stated in equation (4). Consider the setting of equation (4). Ignoring log factors and constants, the confidence bound of the IPS estimator essentially scales as $\sqrt{\frac{\sigma^2\|v\|^2_{A(\bar{\lambda})^{-1}}}{T}} + \frac{\max_{x\in\mathcal{X}} |v^\top A(\bar{\lambda})^{-1}x|}{T}$, while the confidence bound of RIPS essentially scales as $\sqrt{\frac{\sigma^2\|v\|^2_{A(\bar{\lambda})^{-1}}}{T}}$. It can be shown that in the instance in the experiment corresponding to figure 4, the term $\frac{\max_{x\in\mathcal{X}} |v^\top A(\bar{\lambda})^{-1}x|}{T} \approx \frac{d}{T}$ while $\sqrt{\frac{\sigma^2\|v\|^2_{A(\bar{\lambda})^{-1}}}{T}} \approx \frac{\sqrt{d}}{\sqrt{T}}$. Thus, the first term dominates by a polynomial factor in the dimension until $T \ge d$, and the experiment shows that indeed the IPS estimator has larger deviations than RIPS, as suggested by the above upper bounds.

2.3. Experimental Design optimization in an RKHS

We now discuss how to actually compute an allocation in a potentially infinite dimensional RKHS $\mathcal{H}$. The following lemma will be helpful and is proved in the appendix.

Lemma 1. *If* $A_\lambda = \sum_{x\in\mathcal{X}} \lambda_x \phi(x)\phi(x)^\top$ *then for* $a, b \in \mathcal{H}$

$$a^\top(A_\lambda + \gamma I)^{-1}b = \frac{1}{\gamma}a^\top b - \frac{1}{\gamma}k_\lambda(a)^\top(K_\lambda + \gamma I_{|\mathcal{X}|})^{-1}k_\lambda(b)$$

with $k_\lambda(\cdot) \in \mathbb{R}^{|\mathcal{X}|}$ *so that for any* $c \in \mathcal{H}$, $[k_\lambda(c)]_i = \sqrt{\lambda_i}\phi(x_i)^\top c$*, and* $K_\lambda \in \mathbb{R}^{|\mathcal{X}|\times|\mathcal{X}|}$ *so that*

$$[K_\lambda]_{i,j} = \sqrt{\lambda_i}\sqrt{\lambda_j}\phi(x_j)^\top\phi(x_j) =: \sqrt{\lambda_i}\sqrt{\lambda_j}k(x_i, y_i).$$

For $x \in \mathcal{X}$, $[k_\lambda(x)]_i = \sqrt{\lambda_i}\phi(x_i)^\top\phi(x) = \sqrt{\lambda_i}k(x_i, x)$.

If we call $f(\lambda)$ the argument of Equation 6 in Figure 1, and $\bar{v} \in \arg\max_{v\in\mathcal{V}} v^\top\left(\sum_{x\in\mathcal{X}} \lambda_x\phi(x)\phi(x)^\top + \gamma I\right)^{-1} v$ then the computation of the gradient of $\lambda \mapsto f(\lambda)$ equals

$$[\nabla_\lambda f(\lambda)]_i = -\left(\bar{v}^\top(\sum_{x\in\mathcal{X}} \lambda_x\phi(x)\phi(x)^\top + \gamma I)^{-1}\phi(x_i)\right)^2.$$

Importantly, in this work $\mathcal{V}$ will always be a linear combination of $\{\phi(x)\}_{x\in\mathcal{X}}$ (e.g. $\mathcal{V} = \mathcal{X} - \mathcal{X}$), thus the last quantity can be computed only using kernel evaluations thanks to Lemma 1. We use first order optimization methods to minimize $\lambda \mapsto f(\lambda)$ since it is convex.

2.4. Project-Then-Round (PTR) for RKHS designs

To the best of our knowledge, the RIPS procedure of Figure 1 is novel and should be benchmarked. To design a baseline, we take inspiration from previous works on experimental design in an RKHS. For instance, (Alaoui & Mahoney, 2015) employ a sampling distribution related to statistical leverage scores to construct a sketch of the kernel matrix using a Nystrom approximation. The objective in that problem is closest to V-optimal design which aims to minimize the sum-squared error $\sum_{x\in\mathcal{X}} \mathbb{E}[\langle x, \widehat{\theta} - \theta_*\rangle^2]$ (note, our work is concerned with G-optimal-like objectives, or worst-case error over $\mathcal{X}$). The Nystrom approximation to the kernel matrix effectively projects the problem to a low dimensional sub-space where finite-dimensional rounding techniques like those reviewed in Section 1.2 can be applied. (Bach, 2015) also relies on a sampling distribution to approximate integrals using kernels with an objective similar to V optimal.

We describe in Algorithm 2.4 the baseline procedure we call Project-Then-Round (PTR), that employs the finite rounding technique of (Allen-Zhu et al., 2017) described in Section 1.2. This procedure enjoys the following guarantees.

Theorem 2. *Consider the procedure of Algorithm 2.4. If the number of measurements τ satisfies $\tau = \Omega(\widetilde{d}(\gamma, \lambda))$, then*

$$\max_{v\in\mathcal{V}} \|v\|^2_{\left(\sum_{i=1}^{\tau} \phi(\widetilde{x}_i)\phi(\widetilde{x}_i)^\top + \tau\gamma I\right)^{-1}} \leq \max(2, 1+\epsilon) \max_{v\in\mathcal{V}} \|v\|^2_{\left(\sum_{x\in\mathcal{X}} \tau\lambda_x \phi(x)\phi(x)^\top + \tau\gamma I\right)^{-1}}.$$

where $\widetilde{d}(\gamma, \lambda)$ is defined in the algorithm.

We refer the reader to the appendix for the proof of Theorem 2. This procedure performs rounding in a finite dimensional subspace which is a projection of the initial feature space of potentially infinite dimension. With Theorem 2 one can obtain a guarantee similar to that of Theorem 1 up to a constant whenever $\tau = \Omega(\widetilde{d}(\lambda, \gamma))$. Though this effective dimension is rarely the dominating factor in analyses, it is cumbersome to keep around and bound.

2.5. Empirical evaluation of allocation methods

We briefly describe illustrative experiments (see the supplementary material for more details).

G-optimal design experiment: We generate $x_1, \ldots, x_n$ by sampling $\tilde{x}_i \sim N(0, \Sigma)$ with $\Sigma_{i,i} = 1$ if $i \leq d - 10$, $\Sigma_{i,i} = .1$ if $i > d - 10$ and all other entries of Σ set to 0. Then, we set $x_i = \frac{\tilde{x}_i}{\|\tilde{x}_i\|}$. We use $\theta_* = \frac{1}{\sqrt{d}}\mathbf{1}$. We set $d = 50$ and $n = \frac{d(d+1)}{2}$. We use mirror descent to solve the G-optimal design problem. We compare RIPS with IPS, Caratheodory's algorithm with the ceiling rounding technique (LS Cara), the rounding technique in (Allen-Zhu et al., 2017) (LS Regsel), and the random sampling approach taken in (Rizk et al., 2020) (LS Sampling). Figure 2 depicts the results, and shows that RIPS performs comparably to these other approaches. It also illustrates the shortcomings of the Caratheodory rounding algorithm, which does not return an estimate for $T \leq 1275$, while the other algorithms have already learned nontrivial estimates of θ_* for much smaller values of T.

G-optimal design in an RKHS: We let $\mathcal{X} = \{0, (\frac{1}{m})^2, \ldots, (\frac{m-1}{m})^2, 1\}$ with $m = 500$ and use the RBF kernel $K(x, x') = \exp(-\frac{\|x-x'\|^2}{2\varphi^2})$ with bandwidth parameter $\varphi = 0.025$. Due to this being an infinite dimensional kernel, the ambient dimension for m points is equal to m. We focus on the regime $T < m$ where standard rounding schemes do not apply and compare PTR with regularization γ, $\widehat{\theta}_{RIPS}(\gamma)$, and $\widehat{\theta}_{IPS(\gamma)} := A^{(\gamma)}(\lambda)^{-1}(\frac{1}{T}\sum_{t=1}^{T} x_t y_t)$ where we set $\gamma = 0.005$. Figure 3 depicts the results, showing that PTR(γ) does slightly better than IPS(γ) and RIPS(γ), and that all three algorithms have learned nontrivial estimates of θ_* using hundreds of samples fewer than standard rounding algorithms require to even output an estimate.

RIPS vs. IPS: While IPS has similar performance to RIPS in the two previous experiments, RIPS performs dramatically better in some settings. Let $m \in \mathbb{N}$ and $d = m^2 + m$. Inspired by combinatorial bandits, we consider a setting where the measurement vectors $\mathcal{X} = \{e_1, \ldots, e_d\}$ consist of the standard basis vectors, $\theta_* = -\mathbf{1}$, and the performance metric for an estimator $\widehat{\theta}$ is $\mathbb{E}\sup_{i\in[m^2]} |v_i^\top(\widehat{\theta} - \theta_*)|$ where $v_i = \sum_{j=1}^{m} e_j + e_{i+m}$. We compare the performance of IPS against RIPS for $m \in \{12, 14, 16\}$ and estimate the expected maximum deviation at $T = 4m$. Figure 4 shows that as m grows, the performance of IPS degrades relative to RIPS, reflecting that IPS has large deviations in comparison to our proposed estimator RIPS.

3. Algorithms for Kernelized Bandits

We now leverage our proposed RIPS estimator of Algorithm 1 for the kernel bandits problem in an elimination style algorithm as introduced in Section 1.1. In this section we provide different algorithms to solve the regret minimization and pure exploration problems. This section illustrates the benefits of using our RIPS estimator. In particular, the estimator enables us to design a regret minimization algorithm that trivially supports batching while enjoying state of

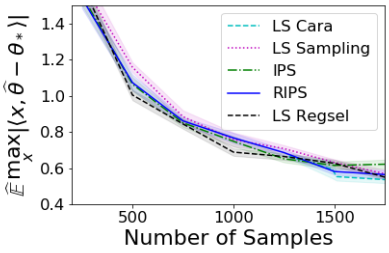

Figure 2. G-Optimal Design Experiment

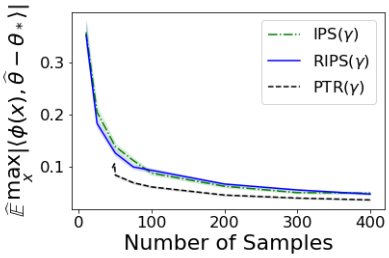

Figure 3. Kernel Experiment

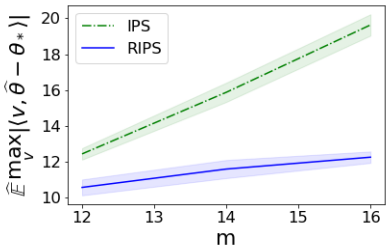

Figure 4. RIPS vs IPS Experiment

Algorithm 2 PTR for Experimental Designs in an RKHS

Input: Finite sets $\mathcal{X} \subset \mathbb{R}^d$ and $\mathcal{V} \subset \mathcal{H}$, feature map $\phi : \mathbb{R}^d \to \mathcal{H}$, regularization $\gamma > 0$
Fix any $\lambda \in \triangle_{\mathcal{X}}$.
Compute $[K]_{i,j} = k(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle_{\mathcal{H}}$ the kernel matrix of the set of points $\mathcal{X} = \{x_1, \ldots, x_n\}$.
Consider a decomposition $K = \widehat{\Phi}\widehat{\Phi}^\top$ with $\widehat{\Phi} \in \mathbb{R}^{n \times n}$ such that the rows of $\widehat{\Phi}$ called $\widehat{\phi}(x_i) \in \mathbb{R}^n$ are used to compute $\widehat{A}(\lambda) = \sum_{i=1}^n \lambda_i \widehat{\phi}(x_i)\widehat{\phi}(x_i)^\top$.
Diagonalize $\widehat{A}(\lambda)$ as $\widehat{A}(\lambda) = VDV^\top$ with D diagonal matrix with coefficients $(d_1 \geq d_2 \geq \ldots \geq d_n)$.
Define the *effective dimension* as

$$\widetilde{d}(\lambda, \gamma) = \max\{i \in [n] : d_i \geq \gamma\}.$$

Choose $k = \widetilde{d}(\gamma, \lambda) \in [n]$ and denote V_k as the top k eigenvectors of $\widehat{A}(\lambda)$.
Compute the projections $V_k^\top \widehat{\phi}(x_1), \ldots, V_k^\top \widehat{\phi}(x_n) \in \mathbb{R}^k$
Use the rounding procedure of (Allen-Zhu et al., 2017) to obtain the desired sparse allocation $\{\widetilde{x}_i\}_{i=1}^\tau$.
Return: $\{\widetilde{x}_i\}_{i=1}^\tau$

the art performance in the well-specified setting. In addition, this same algorithm is robust to model misspecification, suffering only linear regret with respect to that approximation error without any prior knowledge on this error (guarantees that, to the best of our knowledge, our novel). Last but not least, applying our RIPS estimator to pure exploration tasks leads to the first best arm identification provably robust to misspecification.

3.1. RIPS for Regret minimization

As introduced in Section 1, our objective is to develop an algorithm that minimizes regret under the general stochastic and misspecified setting (Assumptions 1-3). Specifically, when pulling arm $x \in \mathcal{X}$ at time t we observe a random variable $\mu_x + \xi_t$ where ξ_t is independent, mean-zero noise with variance σ^2. We assume there exists a $\theta_* \in \mathcal{H}$ and known feature map $\phi : \mathcal{X} \to \mathcal{H}$ such that $\max_{x \in \mathcal{X}} |\mu_x - \langle \theta_*, \phi(x) \rangle| \leq h$ where $h \geq 0$ is unknown to the learner. That is, μ_x is well-approximated by the linear function $\langle \theta_*, \phi(x) \rangle$ but may deviate from it by an amount $h \geq 0$. Because of model misspecification in the case when $h > 0$, we should not hope to obtain sub-linear regret if we seek a regret bound that grows only logarithmically in $|\mathcal{X}|$ and polynomial in d.

Algorithm 3 is a phased elimination strategy where at each round a (regularized) G-optimal design is performed to minimize the variances of the estimates of all the arms and then arms are discarded if their sub-optimality gap is deemed too large (under the assumed linear model). Due to model misspecification, we should only expect this approach to work until hitting a kind of noise floor defined by the level of misspecification h, as suggested from the guarantee from Theorem 1. The algorithm is a combination of our RIPS estimator for the RKHS setting and the robust algorithm of (Lattimore et al., 2020).

Theorem 3. *With probability at least* $1 - \delta$*, the regret of Algorithm 3 satisfies*

$$\sum_{t=1}^T \mu_x - \mu_{x_t} \lesssim c_1 \log(|\mathcal{X}|/\delta) + \sqrt{\max_{\mathcal{V} \subset \mathcal{X}} f(\mathcal{V}, \gamma)} \times \qquad (7)$$

$$\times \left(T(h + \sqrt{\gamma}\|\theta_*\|) + \sqrt{c_0^2(\sigma^2 + B^2) T \log(|\mathcal{X}| \log(T)/\delta)} \right)$$

where $f(\mathcal{V}, \gamma) = \inf_{\lambda \in \triangle_{\mathcal{V}}} \sup_{y \in \mathcal{V}} \|\phi(y)\|^2_{(\sum_{x \in \mathcal{X}} \lambda_x x x^\top + \gamma I)^{-1}}$.

Choosing $\gamma = 1/T$, $\delta = 1/T$ yields an expected regret of

$$\mathbb{E}[\sum_{t=1}^T \mu_{x_*} - \mu_{x_t}] \leq c' \sqrt{\max_{\mathcal{V} \subset \mathcal{X}} f(\mathcal{V}, \tfrac{1}{T})} \left(hT + \sqrt{\log(|\mathcal{X}|T) T} \right)$$

where $c' = O(\sqrt{\|\theta_*\|^2 + \sigma^2 + B^2})$. Note that the hT term due to model misspecification is comparable to the one in (Lattimore et al., 2020). Prior works such as (Srinivas et al., 2009; Valko et al., 2013) have demonstrated expected regret bounds in the well-specified ($h = 0$) setting that scale like $\sqrt{\gamma_T T \log(|\mathcal{X}|)}$ where

$$\gamma_T := \max_{\lambda \in \triangle_{\mathcal{X}}} \log\det(TA^{(0)}(\lambda) + \gamma). \qquad (8)$$

Algorithm 3 RIPS for Regret Minimization

Input: Finite sets $\mathcal{X} \subset \mathbb{R}^d$ ($|\mathcal{X}| = n$), feature map ϕ, confidence level $\delta \in (0,1)$, regularization γ, sub-Gaussian parameter σ, bound on maximum reward B.
Set $\mathcal{X}_1 \leftarrow \mathcal{X}, \ell \leftarrow 1$
while $|\mathcal{X}_\ell| > 1$ **do**
Let $\lambda_\ell \in \triangle_\mathcal{X}$ be a minimizer of $f(\lambda; \mathcal{X}_\ell, \gamma)$ where

$$f(\mathcal{V}, \gamma) = \inf_{\lambda \in \triangle_\mathcal{V}} f(\lambda; \mathcal{V}, \gamma) = \inf_{\lambda \in \triangle_\mathcal{V}} \max_{y \in \mathcal{V}} \|\phi(y)\|^2_{(\sum_{y \in \mathcal{V}} \lambda_y \phi(y)\phi(y)^\top + \gamma I)^{-1}}$$

Set $\epsilon_\ell \leftarrow 2^{-\ell}$, $q_\ell^{(1)} \leftarrow c_1 \log(|\mathcal{X}|/\delta)$
Set $q_\ell^{(2)} \leftarrow c_0^2 (B^2 + \sigma^2) \epsilon_\ell^{-2} f(\mathcal{X}_\ell, \gamma) \log(4\ell^2 |\mathcal{X}|/\delta)$
Set $\tau_\ell \leftarrow \left\lceil \max\left\{ q_\ell^{(1)}, q_\ell^{(2)} \right\} \right\rceil$
Use Algorithm 1 with sets $\mathcal{X}_\ell$, $\mathcal{V}_\ell = \phi \circ \mathcal{X}_\ell$, sampling τ_ℓ measurements $x_1, \ldots, x_{\tau_\ell}$ to get $\{W^{(v)}\}_{v \in \mathcal{V}_\ell}$.
Set $\widehat{\theta}_\ell := \arg\min_\theta \max_{v \in \mathcal{V}_\ell} \frac{|\langle \theta, v \rangle - W^{(v)}|}{\|v\|_{(\sum_{x \in \mathcal{X}} \lambda_{\ell,x} \phi(x)\phi(x)^\top + \gamma I)^{-1}}}$
Update active set:

$$\mathcal{X}_{\ell+1} = \left\{ x \in \mathcal{X}_\ell, \max_{x' \in \mathcal{X}_\ell} \langle \phi(x') - \phi(x), \widehat{\theta}_\ell \rangle < 4\epsilon_\ell \right\}$$

$\ell \leftarrow \ell + 1$
end while
Play unique element of $\mathcal{X}_\ell$ indefinitely.

Algorithm 4 RIPS for Pure Exploration

Input: Finite sets $\mathcal{X} \subset \mathbb{R}^d$, $\mathcal{Z} \subset \mathbb{R}^d$, feature map ϕ, confidence level $\delta \in (0,1)$, regularization γ, sub-Gaussian parameter σ, bound on maximum reward B, bound on the misspecification noise h.
Let $\mathcal{Z}_1 \leftarrow \mathcal{Z}, \ell \leftarrow 1$
while $|\mathcal{Z}_\ell| > 1$ **do**
Let $\lambda_\ell \in \triangle_\mathcal{X}$ be a minimizer of $f(\lambda; \mathcal{Z}_\ell; \gamma)$ where

$$f(\mathcal{V}; \gamma) = \inf_{\lambda \in \triangle_\mathcal{X}} f(\lambda; \mathcal{V}; \gamma) = \inf_{\lambda \in \triangle_\mathcal{X}} \max_{v, v' \in \mathcal{V}} \|\phi(v) - \phi(v')\|^2_{(\sum_{x \in \mathcal{X}} \lambda_x \phi(x)\phi(x)^\top + \gamma I)^{-1}}$$

Set $\epsilon_\ell \leftarrow 2^{-\ell}$, $q_\ell^{(1)} \leftarrow c_1 \log(|\mathcal{Z}|/\delta)$
Set $q_\ell^{(2)} \leftarrow c_0^2 \epsilon_\ell^{-2} f(\mathcal{Z}_\ell; \gamma)(B^2 + \sigma^2) \log(2\ell^2 |\mathcal{Z}|^2/\delta)$
Set $\tau_\ell \leftarrow \left\lceil \max\left\{ q_\ell^{(1)}, q_\ell^{(2)} \right\} \right\rceil$
Use Algorithm 1 with sets $\mathcal{X}$, $\mathcal{V}_\ell = \phi \circ \mathcal{Z}_\ell - \phi \circ \mathcal{Z}_\ell$, sampling τ_ℓ measurements $x_1, \ldots, x_{\tau_\ell}$ to get $\{W^{(v)}\}_{v \in \mathcal{V}_\ell}$.
Set $\widehat{\theta}_\ell := \arg\min_\theta \max_{v \in \mathcal{V}_\ell} \frac{|\langle \theta, v \rangle - W^{(v)}|}{\|v\|_{(\sum_{x \in \mathcal{X}} \lambda_{\ell,x} \phi(x)\phi(x)^\top + \gamma I)^{-1}}}$
$\mathcal{Z}_{\ell+1} = \left\{ z \in \mathcal{Z}_\ell : \max_{z' \in \mathcal{Z}_\ell} \langle \phi(z') - \phi(z), \widehat{\theta}_\ell \rangle \le 2\epsilon_\ell \right\}$
$\ell \leftarrow \ell + 1$
end while
Output: $\mathcal{Z}_\ell$

where $A^{(0)}(\lambda)$ is defined as in (5). The following lemma shows that our own regret bound is never worse than these results.

Lemma 2. *Let γ_T be defined as in* (8)*. Then*

$$\max_{\mathcal{V} \subset \mathcal{X}} f(\mathcal{V}, \tfrac{\gamma}{T}) = \max_{\mathcal{V} \subset \mathcal{X}} \inf_{\lambda \in \triangle_\mathcal{V}} \sup_{y \in \mathcal{V}} \|\phi(y)\|^2_{A^{(\gamma/T)}(\lambda)^{-1}} \le \frac{3}{2}\gamma_T \,.$$

The quantity $f(\mathcal{X}, \gamma)$ can also be bounded by a more interpretable form:

Lemma 3. *If $f(\mathcal{X}, \gamma) = \inf_{\lambda \in \triangle_\mathcal{X}} \max_{y \in \mathcal{X}} \|\phi(y)\|^2_{(A(\lambda) + \gamma I)^{-1}}$ then*

$$\begin{aligned} f(\mathcal{X}, \gamma) &\le \mathrm{Trace}\left(A(\lambda_D^*)(A(\lambda_D^*) + \gamma I)^{-1} \right) \\ &= \mathrm{Trace}\left(K_{\lambda_D^*}(K_{\lambda_D^*} + \gamma I)^{-1} \right) \end{aligned}$$

where $\lambda_D^* \in \arg\max_{\lambda \in \triangle_\mathcal{X}} \log\det\left(A^{(\gamma)}(\lambda)\right)$.

Notably, the RHS of Lemma 3 is the notion of effective dimension that appears in (Alaoui & Mahoney, 2015; Derezinski et al., 2020).

3.2. RIPS for Pure Exploration

We consider a slight generalization of the pure exploration setting introduced in Section 1. Fix finite sets $\mathcal{X} \subset \mathbb{R}^d$ and $\mathcal{Z} \subset \mathbb{R}^d$. We may have $\mathcal{X} = \mathcal{Z}$ but there are interesting cases in which $\mathcal{X} \neq \mathcal{Z}$ including combinatorial bandits and recommendation tasks (Fiez et al., 2019). We say a $z \in \mathcal{Z}$ is ϵ-good if $\mu_z \ge \max_{z' \in \mathcal{Z}} \mu_{z'} - \epsilon$. In the pure exploration game, for $\epsilon > 0$ and $\delta \in (0,1)$ the player seeks to identify an ϵ-good arm by taking as few measurements in $\mathcal{X}$ as possible. Just as in regret minimization games, we assume that when the player at time t plays $x_t \in \mathcal{X}$ she observes $y_t = \mu_{x_t} + \xi_t$ where ξ_t is independent mean-zero noise with variance σ^2. Finally, we assume the existence of a $\theta_* \in \mathcal{H}$ such that

$$\max\left\{ \max_{z \in \mathcal{Z}} |\mu_z - \langle \theta_*, \phi(z) \rangle|, \max_{x \in \mathcal{X}} |\mu_x - \langle \theta_*, \phi(x) \rangle| \right\} \le h$$

for some $h \ge 0$ that is *unknown* to the player.

Consider the elimination style algorithm of Algorithm 4. The algorithm is a combination of our RIPS procedure and the algorithm of (Fiez et al., 2019). While the algorithm is inspired by (Fiez et al., 2019), their analysis only holds in the well-specified setting ($h = 0$), hence a new proof technique was necessary to achieve the following result for general $h \ge 0$.

Theorem 4. *With $z_* \in \arg\max_{z\in\mathcal{Z}}\langle z, \theta_*\rangle$, fix any $\epsilon \geq \bar{\epsilon}$ where*

$$\bar{\epsilon} = 8\min\{\epsilon \geq 0 : 4(\sqrt{\gamma}\|\theta_*\|_2 + h)(2+\sqrt{g(\epsilon)}) \leq \epsilon\},$$
$$g(\epsilon) = \inf_{\lambda\in\triangle_\mathcal{X}} \sup_{z\in\mathcal{Z}:\langle\theta_*,\phi(z_*)-\phi(z)\rangle\leq\epsilon} \|\phi(z_*)-\phi(z)\|^2_{A^{(\gamma)}(\lambda)^{-1}}$$

Then with probability at least $1-\delta$, once the algorithm has taken at least τ samples where $\tau = \widetilde{O}(c_1\log(|\mathcal{Z}|/\delta) + \log(\epsilon^{-1})c_0^2(B^2+\sigma^2)\log(|\mathcal{Z}|/\delta)\rho^*(\gamma,\epsilon))$ we have that $\mu_{\widehat{z}} \geq \max_{z'\in\mathcal{Z}} -\epsilon$ where $\widehat{z}$ is any arm in the set $\mathcal{Z}_\ell$ under consideration after τ pulls and

$$\rho^*(\gamma,\epsilon) = \inf_{\lambda\in\Delta_\mathcal{X}} \sup_{z\in\mathcal{Z}} \frac{\|\phi(z_*)-\phi(z)\|^2_{A^{(\gamma)}(\lambda)^{-1}}}{\max\{\epsilon^2, \langle\theta_*, \phi(z_*)-\phi(z)\rangle^2)\}}. \quad (9)$$

Note that if $\mathcal{X} = \mathcal{Z}$ we have

$$\begin{aligned} g(\epsilon) &= \inf_{\lambda\in\triangle_\mathcal{X}} \sup_{z\in\mathcal{X}:\langle\theta_*,\phi(z_*)-\phi(z)\rangle\leq\epsilon} \|\phi(z_*)-\phi(z)\|^2_{A^{(\gamma)}(\lambda)^{-1}} \\ &\leq 4\inf_{\lambda\in\triangle_\mathcal{X}} \sup_{x\in\mathcal{X}} \|\phi(x)\|^2_{A^{(\gamma)}(\lambda)^{-1}} \\ &\leq 4\mathrm{Trace}((A(\lambda^*_D)+\gamma I)^{-1}A(\lambda^*_D)) \end{aligned}$$

where the last line follows from Lemma 3. This means $\bar{\epsilon}$, the limit on how well one can estimate the maximizing arm, satisfies $\bar{\epsilon} \lesssim (\gamma\|\theta_*\|+h)\mathrm{Trace}((A(\lambda^*_D)+\gamma I)^{-1}A(\lambda^*_D))^{1/2}$. Thus, if we seek an ϵ-good arm, we should choose γ to make this right hand side less than ϵ. Note that $\gamma = 0$ and $h = 0$ implies $\bar{\epsilon} = 0$. If $\phi \equiv$ identity so that $\mathcal{H} = \mathbb{R}^d$, $h = 0$, and $\gamma = 0$ then the sample complexity of Theorem 4 is known to be optimal up to log factors to identify the very best arm (assuming it is unique) relative to any δ-correct algorithm over $\theta_* \in \mathbb{R}^d$ (Soare et al., 2014; Fiez et al., 2019).

3.3. Comparing to the alternative baseline procedure

In Section 2.4 we proposed a natural alternative to our RIPS procedure for experimental design in an RKHS. This PTR baseline leveraged the fact that the added regularization $\gamma > 0$ effectively made many directions irrelevant. Thus, it projected the problem to a low dimensional subspace where it could apply any of the standard rounding techniques for finite dimensions described in Section 1.2. The dimension of this subspace, denoted $\widetilde{d}$, scales like the number of eigenvalues of $\sum_{x\in\mathcal{X}}\lambda^*_x\phi(x)\phi(x)^\top$ that are greater than γ where $\lambda^* \in \arg\min_{\lambda\in\triangle_\mathcal{X}}\max_{v\in\mathcal{V}}\|v\|^2_{\left(\sum_{x\in\mathcal{X}}\lambda_x\phi(x)\phi(x)^\top+\gamma I\right)^{-1}}$. Any standard rounding algorithm would then require the number of samples taken from the design to be at least $\widetilde{d}$. Relative to our results, this inflates our regret bound and sample complexity by an additive factor of $\widetilde{d}$ scaled by some problem-dependent log factors. Algebra shows that $\widetilde{d} \leq 2\mathrm{Trace}((A(\lambda^*_D)+\gamma I)^{-1}A(\lambda^*_D))$. Though for regret this is a lower order term, for pure-exploration with $\mathcal{X} \neq \mathcal{Z}$, this term may potentially dominate the sample complexity because it does not capture the interplay between the geometry of $\mathcal{X}$ and $\mathcal{Z}$. Fortunately, our RIPS procedure demonstrates it is unnecessary and avoids it.

4. Related work

There exist excellent surveys of experimental design from both a statistical and computational perspective (Pukelsheim, 2006; Atkinson et al., 2007; Todd, 2016). Our work is particularly interested in the task of converting a continuous design into a discrete allocation of T measurements. We reviewed a number of works in Section 1.2 for completing this task in finite dimensions. To move to an RKHS setting we considered a regularized design objective which is also known as Bayesian experimental design (Chaloner & Verdinelli, 1995; Allen-Zhu et al., 2017; Derezinski et al., 2020). While most Bayesian experimental design works assume a low-dimensional ambient space and use simple rounding, one exception is the work of (Alaoui & Mahoney, 2015) that performs experimental design in an RKHS for a different design objective, which inspired our project-then-round procedure described of Section 2.4. And very recently, (Derezinski et al., 2020) proposed a method of sampling from a determinantal point process (DPP) and showed that they can approximate many continuous experimental design objectives up to a constant factor if $T \gtrsim d_{eff} := \mathrm{Trace}((A(\lambda^*_D)+\gamma I)^{-1}A(\lambda^*_D))$ with λ^*_D defined in Lemma 3. However, according to Table 1 of (Derezinski et al., 2020) the method may not apply to G-optimal-like objectives[1], which is the primary objective of our work. To our knowledge, our proposed RIPS method is novel in that its performance is directly comparable to the continuous design without requiring a minimum number of measurements with some dependence on the (effective) dimension. However, our method does require the number of measurements to exceed $\log(|\mathcal{V}|)$. While we leveraged experimental design techniques for kernel bandits, many prior works were able to obtain regret bounds and pure-exploration results using other methods.

Kernel bandits In the well-specified setting ($h = 0$) (Srinivas et al., 2009) propose a UCB style algorithm (Auer et al., 2002) for the RKHS setting. Independently, (Grünewälder et al., 2010) developed similar methods for minimizing simple regret. (Srinivas et al., 2009) established a regret bound of $\sqrt{T}(\|\theta_*\|\sqrt{\gamma_T}+\gamma_T)$ where γ_T is defined in (8). (Valko et al., 2013) proposed another UCB variant to obtain a regret bound that scales just as $\|\theta_*\|\sqrt{p_T T}$ where p_T is an algorithm-dependent constant that can be upper bounded by γ_T, thus improving (Srinivas et al., 2009). We recall that our own regret bound of Theorem 3 scales no worse than

[1] Our Theorem 2 with the fact $\widetilde{d} \leq 2d_{eff}$ suggests k only needs to be at least $\widetilde{d}$ for G-like objectives, which adds to their table.

$\|\theta_*\|\sqrt{\gamma_T T}$ using Lemma 2, thus matching state of the art. (Chowdhury & Gopalan, 2017) offer improvements in regret over GP-UCB when the action space is infinite. We also note that our algorithm naturally allows batch querying, a property that UCB-like algorithms achieve only through inelegant means (Desautels et al., 2012; Wu & Frazier, 2018).

Misspecified models Our approach to misspecified models draws inspiration from (Lattimore et al., 2020) which addresses linear bandits in finite dimensions. Their regret bound scales quadratically in the ambient dimension due to rounding effects. Our RIPS procedure extends this work to an RKHS. The misspecified model setting is related to the corrupted setting where an adversary can choose to corrupt the observed reward by c_t in each round t. Any algorithms for this adversarial setting can also be used to solve kernelized multi-armed bandit in the misspecified setting with total amount of corruption equal to at most $C_T = \sum_{t=1}^{T} c_t = hT$. Using this reduction, the regret bound for the corrupted setting of (Bogunovic et al., 2020) scales like $C_T\sqrt{\gamma_T T}$. Unfortunately, if we take $C_T = hT$ this bound is vacuous. Whether robust algorithms like (Gupta et al., 2019) can be extended to our kernel bandit setting is an open question. Concurrently, (Lee et al., 2021) independently proposed a very similar estimator and algorithm for the related task of solving adversarial bandits.

Constrained linear bandits If we assumed that $\|\theta_*\|_2 \leq R$ for some explicit, known $R > 0$ then this setting is known as constrained linear bandits, tackled in (Degenne et al., 2020) for the pure-exploration and (Tirinzoni et al., 2020) for the regret setting, respectively. There, a lower bound on the sample complexity of identifying the best arm can be computed. The lower bound is $\inf_{\lambda\in\Delta_\mathcal{X}}\sup_{x'\neq x_*}\inf_{\gamma\geq 0} G^{-1}(\lambda, x, \gamma)$ where

$$G(\lambda, x, \gamma) = \frac{\max\{(x' - x_*)^\top (A(\lambda) + \gamma I)^{-1} A(\lambda)\theta_*, 0\}^2}{2\|x' - x_*\|^2_{(A(\lambda)+\gamma I)^{-1}}} + \frac{\gamma}{2}\left(\|\theta_*\|^2_{(A(\lambda)+\gamma I)^{-1}A(\lambda)} - R^2\right),$$

which is close to our upper bound ρ^* from equation 9. See Corollary 1 for the proof of this lower bound.
(Degenne et al., 2020) propose an algorithm with an asymptotic upper bound in the sense that as $\delta \to 0$, the dominant term matches the lower bound. However, while (Degenne et al., 2020) and (Tirinzoni et al., 2020) are tight asymptotically, they suffer from large sub-optimal dependencies on problem-specific parameters.

5. Conclusion

In this paper, we have brought to the non-parametric learning setting an estimator that relies on continuous designs while enjoying state of the art - theoretical and experimental - guarantees for both the well-specified and the misspecified settings. We leveraged this estimator in a novel elimination style algorithm for kernel bandits. For the most part we have ignored computation. However, the computational cost of the RIPS estimator scales *linearly* in $|\mathcal{V}|$. An interesting avenue of research is designing an estimator that leverages multi-dimensional robust mean estimation that has the same properties as RIPS but has *no* dependence on $|\mathcal{V}|$. Such an estimator would be of considerable interest in problems such as combinatorial bandits where $|\mathcal{V}|$ is potentially exponential in the dimension (e.g., see (Katz-Samuels et al., 2020; Wagenmaker et al., 2021)).

Acknowledgments

The work of RC and KJ is supported in part by grants NSF RI 1907907 and NSF CCF 2007036.

References

Alaoui, A. E. and Mahoney, M. W. Fast randomized kernel methods with statistical guarantees, 2015.

Allen-Zhu, Z., Li, Y., Singh, A., and Wang, Y. Near-optimal design of experiments via regret minimization. In *International Conference on Machine Learning*, pp. 126–135. PMLR, 2017.

Atkinson, A., Donev, A., and Tobias, R. *Optimum experimental designs, with SAS*, volume 34. Oxford University Press, 2007.

Auer, P., Cesa-Bianchi, N., and Fischer, P. Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47:235–256, 05 2002. doi: 10.1023/A:1013689704352.

Bach, F. On the equivalence between kernel quadrature rules and random feature expansions, 2015.

Bogunovic, I., Krause, A., and Scarlett, J. Corruption-tolerant gaussian process bandit optimization, 2020.

Chaloner, K. and Verdinelli, I. Bayesian experimental design: A review. *Statistical Science*, pp. 273–304, 1995.

Chowdhury, S. R. and Gopalan, A. On kernelized multi-armed bandits, 2017.

Degenne, R., Ménard, P., Shang, X., and Valko, M. Gamification of pure exploration for linear bandits. In *International Conference on Machine Learning*, pp. 2432–2442. PMLR, 2020.

Derezinski, M., Liang, F., and Mahoney, M. Bayesian experimental design using regularized determinantal point processes. In *International Conference on Artificial Intelligence and Statistics*, pp. 3197–3207. PMLR, 2020.

Desautels, T., Krause, A., and Burdick, J. Parallelizing exploration-exploitation tradeoffs with gaussian process bandit optimization, 2012.

Fiez, T., Jain, L., Jamieson, K., and Ratliff, L. Sequential experimental design for transductive linear bandits. *NeurIPS*, 2019.

Frazier, P. I. A tutorial on bayesian optimization. *arXiv preprint arXiv:1807.02811*, 2018.

Grünewälder, S., Audibert, J.-Y., Opper, M., and Shawe-Taylor, J. Regret bounds for gaussian process bandit problems. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pp. 273–280. JMLR Workshop and Conference Proceedings, 2010.

Gupta, A., Koren, T., and Talwar, K. Better algorithms for stochastic bandits with adversarial corruptions. In *Conference on Learning Theory*, pp. 1562–1578. PMLR, 2019.

Katz-Samuels, J., Jain, L., Jamieson, K. G., et al. An empirical process approach to the union bound: Practical algorithms for combinatorial and linear bandits. *Advances in Neural Information Processing Systems*, 33, 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, 2020.

Lee, C.-W., Luo, H., Wei, C.-Y., Zhang, M., and Zhang, X. Achieving near instance-optimality and minimax-optimality in stochastic and adversarial linear bandits simultaneously, 2021.

Lugosi, G. and Mendelson, S. Mean estimation and regression under heavy-tailed distributions: A survey. *Foundations of Computational Mathematics*, 19(5):1145–1190, 2019.

Nikolov, A., Singh, M., and Tantipongpipat, U. T. Proportional volume sampling and approximation algorithms for a-optimal design. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 1369–1386. SIAM, 2019.

Pukelsheim, F. *Optimal design of experiments*. SIAM, 2006.

Rizk, G., Colin, I., Thomas, A., and Draief, M. Refined bounds for randomized experimental design. *arXiv preprint arXiv:2012.15726*, 2020.

Soare, M., Lazaric, A., and Munos, R. Best-arm identification in linear bandits. *arXiv preprint arXiv:1409.6110*, 2014.

Srinivas, N., Krause, A., Kakade, S. M., and Seeger, M. Gaussian process optimization in the bandit setting: No regret and experimental design. *arXiv preprint arXiv:0912.3995*, 2009.

Tao, C., Blanco, S., and Zhou, Y. Best arm identification in linear bandits with linear dimension dependency. In *International Conference on Machine Learning*, pp. 4877–4886, 2018.

Tirinzoni, A., Pirotta, M., Restelli, M., and Lazaric, A. An asymptotically optimal primal-dual incremental algorithm for contextual linear bandits. *arXiv preprint arXiv:2010.12247*, 2020.

Todd, M. J. *Minimum-volume ellipsoids: Theory and algorithms*. SIAM, 2016.

Valko, M., Korda, N., Munos, R., Flaounas, I., and Cristianini, N. Finite-time analysis of kernelised contextual bandits, 2013.

Wagenmaker, A., Katz-Samuels, J., and Jamieson, K. Experimental design for regret minimization in linear bandits. In *International Conference on Artificial Intelligence and Statistics*, pp. 3088–3096. PMLR, 2021.

Wu, J. and Frazier, P. I. The parallel knowledge gradient method for batch bayesian optimization, 2018.