Trajectory-Level Experimental Design for Fast Safety Parameter Estimation of Unknown Environments by Autonomous Systems

Aneesh Raghavan ANEESH@KTH.SE
DCS Division, KTH, Royal Institute of Technology, Stockholm

Karl H. Johansson KALLEJ@KTH.SE
DCS Division, KTH, Royal Institute of Technology, Stockholm

Editors: G. Sukhatme, L. Lindemann, S. Tu, A. Wierman, N. Atanasov

Abstract

We consider the problem of exploring an unknown environment to identify safe and unsafe regions, with the objective of minimizing the number of samples required. The safety of each region is parameterized, and these parameters must be estimated. The exploration problem is formulated as maximizing the spectral gap (or equivalently, minimizing the mixing time) of the Markov chain induced by the agent's policy and current parameter estimates. A closed-form solution to the resulting policy optimization problem is derived, leading to an adaptive exploration algorithm in which regions, once labeled as safe or unsafe, are no longer visited. We analyze the sample complexity required to complete the labeling task with high confidence, compare the proposed method against uniform random and Bayesian exploration strategies, and identify sufficient conditions under which the proposed algorithm achieves lower sample complexity.

Keywords: Safety of Autonomous Systems, Parameter Estimation, Experiment Design

1. Introduction

With advancements in autonomy, safety of the agents and other entities in the environment while completing tasks has gained paramount importance. Safe exploration and exploitation have been investigated (Schreiter et al., 2015; Li et al., 2024; Bottero et al., 2022; Sui et al., 2015; Wan et al., 2022) to address this challenge. Safety constraints have been incorporated using different metrics including set-theoretic, distributional, and information theoretic constraints, etc. The resulting policies typically trade off safety and exploration efficiency, ensuring that constraint violations occur with vanishing probability while still guaranteeing sufficient state-space coverage.

Trajectory-oriented Bayesian experimental design has been investigated to enhance performance of parameter estimation problem in medicine and engineering applications (Foster et al., 2019; Huang et al., 2024; Fiez et al., 2019; Tong and Koller, 2000; Weber et al., 2012). Most of the approaches use a combination of frequentist and Bayesian approaches to design policies which maximize the information gathered. Adaptive sampling for fast classification has been investigated in the context of large data in Djouzi et al. (2022); Singh et al. (2017); Shekhar et al. (2021) where sequential sampling algorithms are presented to enhance the learning process. Such policies are typically adaptive and information-directed, concentrating sampling effort in regions that maximize expected information gain or reduce posterior uncertainty most rapidly. Adaptive sampling methods have been used to survey and learn about algal bloom, water quality models, etc, see, for e.g., Zhang and Sukhatme (2007), Stankiewicz et al. (2021), and Fossum et al. (2020).

We consider a parameter estimation problem aimed at minimizing the sample complexity required to identify safe and unsafe regions in an unknown environment. The domain is partitioned into m regions, each associated with an unknown safety parameter taking values in (0,1). An agent

moves between the centers of these regions according to a policy, and upon visiting a region, it receives a binary observation indicating whether the region is safe. These observations are used to form maximum likelihood estimates of the safety parameters, and a region is labeled as safe or unsafe once the estimate meets a prescribed accuracy level. The the objective is to minimize the total number of samples required to complete this labeling task.

This problem is formulated as minimizing the mixing time (or equivalently maximizing the spectral gap) of the induced Markov chain while adapting the stationary distribution to reflect the current difficulty of estimating each region's parameter. A closed-form solution to the optimization problem is presented and an adaptive exploration algorithm is presented. We further derive high-probability bounds on the number of samples required to complete the labeling task. The sample complexity of the proposed algorithm is compared with uniform random exploration and Bayesian exploration.

The outline of the paper is as follows: In Section 2, we formally define the problem to be solved in this paper. In Section 3, we present the solution to the optimization problem and the exploration algorithm. We present a detailed analysis of the sample complexity in Section 4. We conclude with some future directions in Section 5.

2. Problem Formulation

We begin this section by describing the model for the sequential decision making problem and the parameter estimation problem in subsection 2.1. The problems to be solved are formulated in subsection 2.2.

2.1. Model

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the abstract probability space. The domain in which safe and unsafe regions are to be identified is partitioned into m regions and the center of region j is denoted by $\bar{x}_j$. The state of the agent at time instant n denoted by X_n, is a random variable such that $X_n(\omega) \in \mathcal{S},\ \forall n \in \mathbb{N}, \omega \in \Omega$, where $\mathcal{S} = \{\bar{x}_1, \ldots, \bar{x}_m\}$. The observation collected by the agent at state X_n is a random variable Y_n, where $Y_n(\omega) \in \{0, 1\},\ \forall n \in \mathbb{N}, \omega \in \Omega$. The realizations of random variables (X_n, Y_n) are denoted by (x_n, y_n) respectively. The conditional distribution of Y_n given $\{X_j = x_j\}_{j=1}^n, \{Y_j = y_j\}_{j=1}^n$ is defined as,

$$\mathbb{P}(Y_n = 1 | \{X_j = x_j\}_{j=1}^n, \{Y_j = y_j\}_{j=1}^{n-1}) = \mathbb{P}(Y_n = 1 | X_n = \bar{x}_k) = p_k,$$
$$\mathbb{P}(Y_n = 0 | \{X_j = x_j\}_{j=1}^n, \{Y_j = y_j\}_{j=1}^{n-1}) = \mathbb{P}(Y_n = 0 | X_n = \bar{x}_k) = 1 - p_k,$$

where $x_j = \bar{x}_\ell$ for some $\ell \in 1, \ldots, m$ and all $j \in \{1, \ldots, n-1\}$. $\{p_j\}_{j=1}^m$ are unknown and are to be estimated. The space of all admissible policies is defined as,

$$\Pi_{\text{adm}} = \left\{ \pi = \{\pi_n\} \middle| \pi_n : \mathcal{S} \times \{0, 1\} \to \Delta(\mathcal{S}), \pi_n \text{ measurable }, \forall n \geq 1 \right\},$$

where $\Delta(\mathcal{S})$ is the set of distributions on the state space $\mathcal{S}$. The space of stationary policies is a subset such that, $\Pi_{\text{stat}} = \left\{ \pi \in \Pi_{\text{adm}} \middle| \pi_n = \pi_{n+1}, \forall n \geq 1 \right\}$. Let the augmented state be $\widehat{X}_n = (X_n, Y_n),\ \widehat{X}_n \in \mathcal{S} \times \{0, 1\} = \{(\bar{x}_\ell, y) : \ell = 1, \ldots, m,\ y \in \{0, 1\}\}$. Under a fixed policy π, the one-step transition matrix on $\mathcal{S} \times \{0, 1\}$ is

$$P^{(\pi_n)}_{(j,u)\to(\ell,y)} = \Pr(\widehat{X}_{n+1} = (\bar{x}_\ell, y) \mid \widehat{X}_n = (\bar{x}_j, u)) = \pi^\ell_{n,j,u}\, p^v_\ell,$$

where $\pi^{\ell}_{n,j,u} := \pi_n(\bar{x}_j, u)[\bar{x}_\ell]$, $p^y_\ell := \mathbf{1}_{\{y=1\}} p_\ell + \mathbf{1}_{\{y=0\}}(1-p_\ell)$. Given a admissible policy π and a random trajectory $\{X_n, Y_n\}$, define,

$$N_{n,\ell} := \sum_{t=1}^{n} \mathbf{1}_{\{X_t = \bar{x}_\ell\}}, \; S_{n,\ell} := \sum_{t=1}^{n} \mathbf{1}_{\{X_t=\bar{x}_\ell\}} Y_t.$$

The log-likelihood function, $\mathcal{L}(\cdot)$, can be defined as below and an estimate of p_ℓ at iteration n, $\widehat{p}_{n,\ell}$, can be obtained by maximizing the log-likelihood with respect to each p_ℓ as below.

$$\mathcal{L}(\{p_j\}) = \sum_{\ell=1}^{m} \Big[S_{n,\ell} \log p_\ell + (N_{n,\ell} - S_{n,\ell}) \log(1-p_\ell) \Big] \implies \widehat{p}_{n,\ell} = \frac{S_{n,\ell}}{N_{n,\ell}}.$$

Definition 1 *Let $\varepsilon_s \in (0, \frac{1}{4})$ be a given confidence threshold. At iteration n, given $\{\widehat{p}_{n,\ell}\}_{\ell=1}^m$, generated from a admissible policy π, $\texttt{Label}_n(\bar{x}_\ell) = \texttt{SAFE}$ if $\widehat{p}_{n,\ell} > \frac{1}{2}$ and $\widehat{p}_{n,\ell}(1-\widehat{p}_{n,\ell}) < \varepsilon_s$. $\texttt{Label}_n(\bar{x}_\ell) = \texttt{UNSAFE}$ if $\widehat{p}_{n,\ell} < \frac{1}{2}$ and $\widehat{p}_{n,\ell}(1-\widehat{p}_{n,\ell}) < \varepsilon_s$. Once a state is labeled as $\texttt{SAFE}$ or $\texttt{UNSAFE}$, the agent does not visit that state, i.e., the state is deactivated. Let the random number of active states at iteration n be denoted by $\mathcal{S}_n$. At iteration n, if $\texttt{Label}_n(\bar{x}_\ell) \in \{\texttt{SAFE}, \texttt{UNSAFE}\}$, then $\mathcal{S}_n = \mathcal{S}_{n-1} \sim \{\bar{x}_\ell\}$, where $\mathcal{S}_0 = \mathcal{S}$. The iteration at which the kth elimination happens is denoted by $T^{(k)}$, where $k \leq m$. The time period between two eliminations is defined as a phase, i.e., phase k is the interval $[T^{(k-1)}, T^{(k)})$. Let $\mathcal{I}^{(k)}$ be the set of indices of active states for phase k.*

When the true parameters are close to 0 or 1, the above labeling using the confidence threshold ensures that the estimated labels are correct with high confidence. The space of policies is restricted as follows for the problem formulation,

$$\Pi_{\text{opt}} = \Big\{ \pi \in \Pi \Big| \pi_n = \pi^{(k)}, n \in [T^{(k-1)}, T^{(k)}), \{\pi^{(k)}\} \subset \Pi_{\text{stat}}, P^{(k)} \text{ is irreducible and aperiodic} \Big\},$$

where $P^{(k)}$ corresponds to the transition matrix of policy $\pi^{(k)}$. Thus, the joint process, $\{\widehat{X}_n\}$, when driven by a policy $\pi \in \Pi_{\text{opt}}$, is a piecewise homogeneous Markov chain. Further, the process $\{\widehat{X}_n\}$ is ergodic over any phase k and admits a unique stationary distribution $\mu^{(k)}$ satisfying $\mu^{(k)} = \mu^{(k)^\top} P^{(k)}$, (Puterman, 1994; Norris, 1997; Meyn and Tweedie, 2009).

2.2. Problem Description

For a time-homogeneous, irreducible, and aperiodic MDP, $\{\widetilde{X}_n\}$, on a finite state space, driven by stationary policy π, concentration bounds have been found as function of its mixing time $\tau_{\text{mix}}(\pi)$, (Paulin, 2015; Lezaud, 1998; Glynn and Ormoneit, 2002). For an additive functional of the MDP, f, where f is absolutely bounded, a typical concentration bound is of the form,

$$\mathbb{P}\left(\left| \frac{1}{n} \sum_{j=1}^{n} f(\widetilde{X}_j) - \mathbb{E}_\pi[f] \right| \leq B \sqrt{\frac{\tau_{\text{mix}}(\pi) \ln(2/\epsilon)}{c\, n}} \right) \geq 1-\epsilon, \; \epsilon > 0.$$

The convergence of $\dfrac{f_n(\widetilde{X}_n)}{n}$ to $\mathbb{E}_\pi[f]$ follows from the ergodicity of $\{\widetilde{X}_n\}$. In particular, to guarantee $\mathbb{P}(|\frac{1}{n}\sum_{j=1}^n f(\widetilde{X}_j) - \mathbb{E}_\mu[f]| \leq \epsilon)$ with probability at least $1-\delta$, it suffices that $n \gtrsim$

$\tau_{\text{mix}}(\pi)\,\frac{B^2}{\epsilon^2}\,\ln\frac{2}{\delta}$. Bounds for functions like log-likelihood functions which are absolutely bounded have been obtained in (Jedra et al., 2023; Jedra and Proutiere, 2019). The concentration bound as a function of spectral gap has been investigated in (Fan et al., 2021). Utilizing this version of the bound, the experiment design problem to reduce the sample complexity of the MLE estimators is formulated as spectral gap maximization problem for each phase as follows:

Problem 1 *Define* $\beta_\ell^{(k)} := \min\left\{|p_{T^{(k-1)},\ell} - \frac{1}{2}|,\ \frac{\varepsilon_s}{2}\right\}$ *as the gap for center* ℓ *at the beginning of phase* k. *Let the transition matrix for the Markov chain* $\{X_n\}$ *be defined as* $P^{(k)}_{j\to\ell} = (1-\widehat{p}_{T^{(k-1)},j})\,\pi^\ell_{j,0} + \widehat{p}_{T^{(k-1)},j}\,\pi^\ell_{j,1},\ j,\ell\in\mathcal{I}_n$, *where* $\mathcal{I}_n$ *corresponds to the indices of the active states in* $\mathcal{S}_n$. *Let the desired stationary distribution at iteration* n *be* $\mu_{n,\pi}\in\Delta(\mathcal{S}_n)$ *be* $\mu_\ell^{(k)} = \frac{\frac{1}{\beta_\ell^{(k)2}}}{\sum_{\ell\in\mathcal{I}_n}\frac{1}{\beta_\ell^{(k)2}}}$. *The experiment design problem is to maximize the spectral gap of* $P^{(k)}$ *subject to following constraints:*

$$\max_{\pi^{(k)}\in\Pi_{\text{stat}}}\ \gamma^{(k)}(\pi^{(k)}) := 1-\lambda_2\big(P^{(k)}\big),$$
$$\text{subject to } \pi^\ell_{j,1},\pi^\ell_{j,0}\geq\alpha,\ \sum_{\ell\in\mathcal{I}_n}\pi^\ell_{j,1}=1,\ \sum_{\ell\in\mathcal{I}_n}\pi^\ell_{j,0}=1,\ j\in\mathcal{I}_n\ \mu^{(k)}=\mu^{(k)^\top}P^{(k)}.$$

where $\alpha>0$. *This to enforce that the resulting chain is irreducible and aperiodic for the phase* k.

The constraint $\mu^{(k)} = \mu^{(k)^\top}P^{(k)}$ in the above problem enforces the stationary distribution to be inversely proportional to the difficulty of estimating the safety parameter of the active states.

Problem 2 *Consider the algorithm implementing the policy,* $\pi^\star=\{\pi^{(k),\star}\}\in\Pi_{\text{opt}}$, *where* $\pi^{(k),\star}$ *is obtained by solving Problem 1, for completing the labeling task. Given* $\delta>0$, *what is the number of the samples needed by the algorithm to correctly label all centers,* $\{\bar{x}_j\}_{j=1}^m$, *as* `SAFE` *or* `UNSAFE` *with probability at least* $1-\delta$.

3. Algorithm

In this section, first, we present the solution to Problem 1. Then, we present the algorithm executing the policy $\pi^\star$ and completing the labeling task.

Proposition 2 *Let* $m_n = |\mathcal{I}_n|$ *and let* α *be such that* $\alpha\leq\min\left\{\min_{\ell\in\mathcal{I}_n}\mu_\ell^{(k)},\ \frac{1}{m_n}\right\}$. *Then* $\exists\pi^{(k),\star}=\{\pi^\ell_{j,0},\pi^\ell_{j,1}\}$ *such that for every* $j,\ell\in\mathcal{I}_n$ *the following equations are satisfied,*

$$(1-\widehat{p}_{T^{(k-1)},j})\,\pi^\ell_{j,0} + \widehat{p}_{T^{(k-1)},j}\,\pi^\ell_{j,1} = \frac{\frac{1}{\beta_\ell^{(k)2}}}{\sum_{\ell\in\mathcal{I}_n}\frac{1}{\beta_\ell^{(k)2}}},\ \pi^\ell_{j,y}\geq\alpha,\ \sum_{\ell\in\mathcal{I}_n}\pi^\ell_{j,y}=1, y=0,1. \tag{1}$$

Further, $\pi^{(k),\star}$ *solves Problem 1.*

Proof Since $P^{(k)}$ is a stochastic matrix, the maximum eigenvalue of $P^{(k)}$ is 1. By choosing all the rows of $P^{(k)}$ to be identical, the rank of $P^{(k)}$ is 1 and the spectral gap is 1, i.e., the maximum possible value. This implies that every row of $P^{(k)}$ equals $\mu^{(k)^\top}$. The conditions satisfied by $\pi^{(k),\star}$ are

Algorithm 1 AMSE Algorithm: Adaptive Mixing Time Minimization Based Exploration Algorithm for the Identification of `SAFE` and `UNSAFE` Regions

Require: Initial safety estimates $\widehat{p}_{0,\ell} = 1/2$ for all centers ℓ, ε_s, $k = 0$.

while $\mathcal{S}_n \neq \varnothing$ **do**

  if $(\widehat{p}_{n,\ell})(1-\widehat{p}_{n,\ell}) < \varepsilon_s$ for any ℓ or $n = 0$ **then**

    If $\widehat{p}_{n,\ell} > \dfrac{1}{2}$, $\texttt{Label}_n(\bar{x}_\ell) = \texttt{SAFE}$ ELSE $\widehat{p}_{n,\ell} < \dfrac{1}{2}$, $\texttt{Label}_n(\bar{x}_\ell) = \texttt{UNSAFE}$.

    $k \leftarrow k+1, T^{(k-1)} \leftarrow n$, Update $\mathcal{S}_n, \mathcal{I}_n, \mathcal{I}^{(k)}$.

    Compute optimal exploration policy $\pi^{(k),\star}$.

    1. Initialize $x_j^\ell := \alpha$ for all $\ell \in \mathcal{I}_n$. Remaining mass $R := 1 - m\alpha$.

    2. For each ℓ, capacity $c_j^\ell := u_j^\ell - \alpha$, where $u_j^\ell = \frac{\mu_\ell^{(k)} - \widehat{p}_{T^{(k-1)},j}\alpha}{(1-\widehat{p}_{T^{(k-1)},j})}$

    3. Distribute remaining mass R across coordinates x_j^ℓ (e.g. greedily), ensuring $x_j^\ell \leq u_j^\ell$ and $\sum_\ell x_j^\ell = 1$.

    4. $\pi_{j,0}^{(k),\ell} = x_j^\ell$, $\pi_{j,1}^{(k),\ell} = (\mu_\ell^{(k)} - (1-\widehat{p}_{T^{(k-1)},j})x_j^\ell)/\widehat{p}_{T^{(k-1)},j}$, $\pi^{(k),\star} = \{\pi_{j,0}^\ell, \pi_{j,1}^\ell\}$.

  end if

  Implement the resulting policy $\pi^{(k),\star}$. $n \leftarrow n+1$, Gather data point (X_n, Y_n). Update estimates $\widehat{p}_{n,\ell}$ for all ℓ.

end while

return Final set of labeled safe and unsafe centers.

summarized in Equation (1). Now, we derive the conditions on α which ensures that the equations and the inequality in (1) are satisfied. $\pi_{j,1}^\ell$ can be expressed in terms of $\pi_{j,0}^\ell$, taking the sum over ℓ, and using $\sum_\ell \mu_\ell^{(k)} = 1$ gives $\pi_{j,1}^\ell = \frac{\mu_\ell^{(k)} - (1-\widehat{p}_{T^{(k-1)},j})\,\pi_{j,0}^\ell}{\widehat{p}_{T^{(k-1)},j}}$, $\sum_\ell \pi_{j,1}^\ell = \frac{1-(1-\widehat{p}_{T^{(k-1)},j})\sum_\ell \pi_{j,0}^\ell}{\widehat{p}_{T^{(k-1)},j}}$. Imposing $\sum_\ell \pi_{j,1}^\ell = 1$ implies $\sum_\ell \pi_{j,0}^\ell = 1$. Hence, it suffices to find the vector $\{\pi_{j,0}^\ell\}, \ell \in \mathcal{I}_n$, satisfying $\pi_{j,0}^\ell \geq \alpha$, $\sum_\ell \pi_{j,0}^\ell = 1$ and $\pi_{j,1}^\ell$ also satisfy $\pi_{j,1}^\ell \geq \alpha$ for all ℓ. Thus,

$$\frac{\mu_\ell^{(k)} - (1-\widehat{p}_{T^{(k-1)},j})\pi_{j,0}^\ell}{\widehat{p}_{T^{(k-1)},j}} \geq \alpha \quad \Longleftrightarrow \quad \pi_{j,0}^\ell \leq \frac{\mu_\ell^{(k)} - \widehat{p}_{T^{(k-1)},j}\alpha}{(1-\widehat{p}_{T^{(k-1)},j})} =: u_j^\ell.$$

Thus, for each ℓ we must have the box constraint $\alpha \leq \pi_{j,0}^\ell \leq u_j^\ell$ and the simplex constraint $\sum_\ell \pi_{j,0}^\ell = 1$. Therefore, the feasibility problem reduces to checking the existence of $x \in \mathbb{R}^{m_n}$ s.t. $\alpha\mathbf{1} \leq x \leq u$, $\sum_\ell x_j^\ell = 1$. The above feasibility holds if and only if $\alpha \leq u_j^\ell$ for every ℓ (each lower bound does not exceed its upper bound); and $\sum_\ell \alpha \leq 1 \leq \sum_\ell u_j^\ell$ (the total lies between the sum of lower and upper bounds). $\alpha \leq u_j^\ell \iff \mu_\ell^{(k)} \geq \alpha, \ \forall\, \ell$. Moreover, $\sum_\ell \alpha = m_n\alpha$, so $m_n\alpha \leq 1$ is required. Thus, $\alpha \leq \min\{\min_{\ell\in\mathcal{I}_n} q_\ell, \ \frac{1}{m_n}\}$ follows from the former and the condition:

$$1 \leq \sum_\ell u_j^\ell = \frac{1 - m_n\widehat{p}_{T^{(k-1)},j}\alpha}{(1-\widehat{p}_{T^{(k-1)},j})} \iff m_n\widehat{p}_{T^{(k-1)},j}\alpha \leq \widehat{p}_{n,j} \iff m_n\alpha \leq 1.$$

■

4. Analysis

In this section, we study the sample complexity of the proposed algorithm and compare its performance with existing algorithms.

Proposition 3 *Let the true Bernoulli parameter at center ℓ be p_ℓ. Let the gap for center ℓ be $\beta_\ell := \min\left\{|p_\ell - \frac{1}{2}|, \frac{\varepsilon_s}{2}\right\}$. Let $\tau^{(\pi)}$ denote the mixing time, $\mu^{(\pi)}$ denote the stationary distribution, and $\mu^{(\pi)}_{\min} := \min_{\ell\in\mathcal{I}} \mu^\pi_\ell$ its lower bound, under any policy $\pi \in \Pi_{\text{stat}}$. For any $\delta \in (0,1)$, with probability at least $1-\delta$, all states are correctly classified by an algorithm executing policy π after*

$$n \gtrsim \frac{\tau^{(\pi)}}{\mu^{(\pi)}_{\min}} \left(\max_{\ell=1,\dots,m} \frac{1}{\beta_\ell^2}\right) \ln\left(\frac{2m}{\delta}\right).$$

Proof β_ℓ ensures that both the sign relative to $1/2$ and the variance, $p_\ell(1-p_\ell) < \varepsilon_s$, being small is satisfied, i.e., the estimator must be robustly sufficiently close to the true p_ℓ. We use the Markov-chain concentration inequality (subsection 2.2) in its mixing-time form. For a bounded f (here $f(\widehat{X}_n) = \mathbf{1}\{X_n = \bar{x}_\ell\}Y_n$ or $f(\widehat{X}_n) = \mathbf{1}\{X_n = \bar{x}_\ell\}$), the deviation after N_ℓ visits obeys a Hoeffding-type tail with effective sample size $N_\ell/\tau^{(\pi)}$. That is, there exists a constant $c > 0$ such that for $\delta' > 0$, $\mathbb{P}\big(|\hat{p}_\ell - p_\ell| > \epsilon\big) \le 2\exp\Big(-c\,\frac{N_\ell \epsilon^2}{\tau^{(\pi)}}\Big)$. Solving for N_ℓ so that the right-hand side is at most δ' gives $N_\ell \ge \dfrac{\tau^{(\pi)}}{c\,\epsilon^2} \ln\left(\dfrac{2}{\delta'}\right)$. To guarantee correct labeling of center ℓ, it is sufficient that $|\hat{p}_\ell - p_\ell| \le \beta_\ell$. Setting $\epsilon = \beta_\ell$ and applying a union bound across all m centers with total failure probability δ (so $\delta' = \delta/m$) yields $N_\ell \ge \frac{\tau^{(\pi)}}{c\,\beta_\ell^2} \ln\big(\frac{2m}{\delta}\big)$. Visits to state ℓ are random; under stationarity, the fraction of time spent at ℓ is μ^π_ℓ. A conservative deterministic sufficient condition on the total number of samples n is then $n \ge \max_{\ell=1,\dots,m} \frac{N_\ell}{\mu^\pi_\ell}$, which implies $n \ge \frac{1}{\mu^{(\pi)}_{\min}} \max_\ell \frac{\tau_{\max}}{c\,\beta_\ell^2} \ln\big(\frac{2m}{\delta}\big)$. ■

Lemma 4 *Consider the stochastic process $\{X_n, Y_n\}$. In each phase k, a stationary policy $\pi^{(k)}$ is used, inducing a Markov kernel $P^{(k)}$ with mixing time $\tau^{(k)}$ and stationary distribution $\mu^{(k)}$. Assume the following: (A1) Each $P^{(k)}$ is geometrically ergodic with a uniform spectral gap $\gamma > 0$. (A2) Mixing times are uniformly bounded: $\tau^{(k)} \le \tau_{\max}$. (A3) Phase lengths satisfy $T^{(k)} - T^{(k-1)} \gtrsim \tau^{(k)} \log\left(\frac{m}{\delta}\right)$ (burn-in condition). For $\ell \in \mathcal{I}^{(k)}$, define:*

$$\hat{p}^{(k)}_\ell = \frac{1}{N^{(k)}_\ell} \sum_{n=T^{(k-1)}}^{T^{(k)}-1} \mathbf{1}\{X_n = \ell\}Y_n, \text{ where } N^{(k)}_\ell := \sum_{n=T^{(k-1)}}^{T^{(k)}-1} \mathbf{1}\{X_n = \ell\} = \sum_{n\in \textit{phase } k} \mathbf{1}\{X_n = \ell\},$$

$$\hat{p}_\ell = \frac{\sum_k \sum_{n\in \textit{phase } k} \mathbf{1}\{X_n = \ell\}Y_n}{\sum_k N^{(k)}_\ell}, S_\ell = \sum_k S^{(k)}_\ell, S^{(k)}_\ell = \sum_{n\in \textit{phase } k} Z_{n,\ell},\ Z_{n,\ell} = \mathbf{1}\{X_n = \ell\}(Y_n - p_\ell).$$

Then,

$$\mathbb{P}\big(|S_\ell| > x\big) \le 2\exp\left(-\frac{cx^2}{\mathrm{Var}(S_\ell)}\right), \text{ that is, } \mathbb{P}\big(|S_\ell| > x\big) \le 2\exp\left(-\frac{cx^2}{\sum_k N^{(k)}_\ell\, \tau^{(k)}}\right).$$

Proof From the definition of the $\{Z_{n,\ell}\}$ process it follows that $|Z_{n,\ell}| \leq 1,\ \mathbb{E}[Z_{n,\ell} \mid \mathcal{F}_{n-1}] = 0$. Hence $\{Z_{t,\ell}\}$ is a bounded, centered process adapted to the Markov chain. The mixing property of the process follows from the assumptions. Since each phase kernel $P^{(k)}$ has spectral gap $\gamma > 0$ and phase lengths satisfy $T^{(k)} - T^{(k-1)} \gtrsim \tau^{(k)} \log\left(\frac{m}{\delta}\right)$, the full sequence $\{Z_{t,\ell}\}$ is β-mixing with exponential decay $\beta(s) \leq C\exp(-cs/\tau_{\max})$. Justification is as follows. Within each phase, geometric ergodicity implies exponential mixing. The burn-in condition ensures each phase starts close to stationarity. Concatenation of geometrically mixing chains with burn-in preserves exponential β-mixing, (Doukhan, 1995; Bradley, 2005). We now invoke Bernstein inequality for β-mixing sequences. Since $\{Z_{n,\ell}\}$ is a bounded, centered, β-mixing sequence such that, $|Z_{n,\ell}| \leq 1,\ \beta(s) \leq Ce^{-cs}$. Then,

$$\mathbb{P}\left(\sum_{n=1}^{t} Z_{n,\ell} > x\right) \leq 2\exp\left(-\frac{cx^2}{V_t}\right), V_t = \sum_{n=1}^{t} \mathrm{Var}(Z_{n,\ell}) + 2\sum_{n<p} |\mathrm{Cov}(Z_{n,\ell}, Z_{p,\ell})|.$$

The objective is to bound the variance term, $V_t = \mathrm{Var}(S_\ell)$. Within a phase k, from the standard concentration results for geometrically ergodic Markov chains, the concentration bound is as obtained below. The number of visits required per phase is computed by setting $\epsilon = \beta_\ell$.

$$\mathbb{P}\Big(|\hat{p}_\ell^{(k)} - p_\ell| > \epsilon\Big) \leq 2\exp\left(-\frac{c\, N_\ell^{(k)} \epsilon^2}{\tau^{(k)}}\right) \implies N_\ell^{(k)} \gtrsim \tau^{(k)} \cdot \frac{1}{\beta_\ell^2} \log\left(\frac{2m}{\delta}\right).$$

Since $\{Z_{n,\ell}\}$ is a geometrically ergodic Markov chain with mixing time $\tau^{(k)}$, for any f, bounded with $\mathbb{E}_{\mathbb{P}}[f] = 0$, it follows that

$$\mathrm{Var}\left(\sum_{n=T^{k-1}}^{T^k-1} f(Z_{n,\ell})\right) \leq C\,\tau^{(k)} \sum_{n=1}^{t} \mathbb{E}\left[f(Z_{n,\ell})^2\right],$$

where C depends only on mixing constants. The expectation in the R.H.S can be computed as follows.

$$f(Z_{n,\ell})^2 = \mathbf{1}\{X_n = \ell\}(Y_n - p_\ell)^2 \leq \mathbf{1}\{X_n = \ell\} \Longrightarrow \mathbb{E}[f(Z_{n,\ell})^2] = \mathbb{P}(X_n = \ell)\,\mathrm{Var}(Y_n \mid X_n = \ell),$$

$$\Longrightarrow \mathbb{E}[f(Z_{n,\ell})^2] \leq \mathbb{P}(X_n = \ell) \implies \sum_{n=T^{k-1}}^{T^k-1} \mathbb{E}\left[f(Z_{n,\ell})^2\right] \leq \mathbb{E}\left[N_\ell^{(k)}\right] \implies \mathrm{Var}(S_\ell^{(k)}) \lesssim N_\ell^{(k)}\,\tau^{(k)}.$$

The cross covariance can be obtained using β-mixing as $|\mathrm{Cov}(Z_t, Z_s)| \leq C\beta(|t-s|)$ and since $\beta(s) \leq Ce^{-cs/\tau_{\max}}$, if follows that,

$$\sum_{n<p} |\mathrm{Cov}(Z_{n,\ell}, Z_{p,\ell})| \lesssim \sum_{n=1}^{t} \sum_{p=n+1}^{\infty} e^{\frac{-c(s-t)}{\tau_{\max}}} \lesssim \sum_{n=1}^{t} \tau_{\max} \Longrightarrow \sum_{n<p} |\mathrm{Cov}(Z_{n,\ell}, Z_{p,\ell})| \lesssim \sum_{k} N_\ell^{(k)} \tau^{(k)},$$

where the implication follows from the observation that only times with $X_t = \ell$ contribute. Hence, $V_t \lesssim \sum_k N_\ell^{(k)} \tau^{(k)}$. We note that the bound on the cross covariance should ideally be $N_\ell^{(k)} \tau_{\max}$. However due to assumptions (A3), i.e., sufficient gap between phases, we replace $\tau_{\max}$ by $\tau^{(k)}$. Invoking Bernstein's inequality completes the proof. ■

Mixing ensures that long-range dependence is negligible, so variance adds across phases despite local dependence. Thus, the result follows by combining phase-wise Markov concentration (yielding variance $N_\ell^{(k)}\tau^{(k)}$) with mixing-based variance additivity across phases.

Theorem 5 (Phase-wise adaptive elimination with mixing minimization) *Assume the conditions of Lemma 4 hold. For any $\delta \in (0,1)$, with probability at least $1-\delta$, all states are correctly classified after*

$$n \gtrsim \max_{\ell=1,\dots,m} \frac{\tau_{\max}^{(\ell)}}{\bar{\mu}_\ell^{(\ell)}} \cdot \frac{1}{\beta_\ell^2} \log\left(\frac{2m}{\delta}\right),$$

where: $\tau_{\max}^{(\ell)} := \max_{k:\,\ell\in\mathcal{I}^{(k)}} \tau^{(k)}$, $\bar{\mu}_\ell^{(\ell)} := \frac{1}{T_\ell}\sum_{k:\,\ell\in\mathcal{I}^{(k)}} \mu_\ell^{(k)}\left(T^{(k)}-T^{(k-1)}\right)$, is the empirical time-averaged occupancy until elimination, and $T_\ell = \sum_k \left(T^{(k)}-T^{(k-1)}\right)$ is the elimination time of state ℓ.

Proof Define $N_\ell(T_\ell) := \sum_{k:\,\ell\in\mathcal{I}^{(k)}} N_\ell^{(k)}$. Setting $x = \epsilon N_\ell(T_\ell)$ in Lemma 4,

$$\mathbb{P}\big(|S_\ell| > \epsilon N_\ell(T_\ell)\big) \le 2\exp\left(-\frac{c\epsilon^2 N_\ell(T_\ell)^2}{\sum_k N_\ell^{(k)}\tau^{(k)}}\right) \Longrightarrow \mathbb{P}\big(|\hat{p}_\ell - p_\ell| > \epsilon\big) \le 2\exp\left(-\frac{c\epsilon^2 N_\ell(T_\ell)^2}{\sum_k N_\ell^{(k)}\tau^{(k)}}\right),$$

where the implication follows from $\left|\hat{p}_\ell - p_\ell\right| = \frac{|S_\ell|}{N_\ell(T_\ell)}$. Applying Cauchy-Schwarz,

$$\left(\sum_k N_\ell^{(k)}\right)^2 = \left(\sum_k \sqrt{N_\ell^{(k)}\tau^{(k)}} \cdot \sqrt{\frac{N_\ell^{(k)}}{\tau^{(k)}}}\right)^2 \le \left(\sum_k N_\ell^{(k)}\tau^{(k)}\right)\left(\sum_k \frac{N_\ell^{(k)}}{\tau^{(k)}}\right).$$

Rearranging the inequality, that is, dividing both sides by $\sum_k N_\ell^{(k)}\tau^{(k)}$:

$$\frac{N_\ell(T_\ell)^2}{\sum_k N_\ell^{(k)}\tau^{(k)}} \le \sum_k \frac{N_\ell^{(k)}}{\tau^{(k)}} \implies \mathbb{P}\big(|\hat{p}_\ell - p_\ell| > \epsilon\big) \le 2\exp\left(-c\epsilon^2 \sum_k \frac{N_\ell^{(k)}}{\tau^{(k)}}\right).$$

Thus, each phase contributes variance approximately equal to $N_\ell^{(k)}\,\tau^{(k)}$, effective sample size of order $\frac{N_\ell^{(k)}}{\tau^{(k)}}$. Due to weak dependence across phases, contributions add, yielding total information $\sum_k \frac{N_\ell^{(k)}}{\tau^{(k)}}$. To ensure $|\hat{p}_\ell - p_\ell| \le \beta_\ell$ with probability at least $1-\frac{\delta}{m}$, it suffices that:

$$c\,\beta_\ell^2 \sum_k \frac{N_\ell^{(k)}}{\tau^{(k)}} \gtrsim \log\left(\frac{2m}{\delta}\right) \implies \sum_k \frac{N_\ell^{(k)}}{\tau^{(k)}} \gtrsim \frac{1}{\beta_\ell^2}\log\left(\frac{2m}{\delta}\right).$$

$\tau^{(k)} \le \tau_{\max}^{(\ell)} \implies \sum_k \frac{N_\ell^{(k)}}{\tau^{(k)}} \ge \frac{N_\ell(T_\ell)}{\tau_{\max}^{(\ell)}}$. Thus it suffices that $N_\ell(T_\ell) \gtrsim \tau_{\max}^{(\ell)} \cdot \frac{1}{\beta_\ell^2}\log\left(\frac{2m}{\delta}\right)$.

Under assumptions (A1)–(A3) of Lemma 4, each phase kernel $P^{(k)}$ is geometrically ergodic with spectral gap $\gamma > 0$. Therefore, for any phase k and any bounded function f, the additive functional satisfies the inequality,

$$\mathbb{P}\left(\frac{1}{T^{(k)}-T^{(k-1)}}\sum_{t=T^{(k-1)}}^{T^{(k)}-1} f(X_t) - \mathbb{E}_{\mu^{(k)}}[f] > \epsilon\right) \le 2\exp\left(-\frac{c\,(T^{(k)}-T^{(k-1)})\epsilon^2}{\tau^{(k)}}\right).$$

Applying this with $f(x) = \mathbf{1}\{x = \ell\}$ gives $\left| \frac{N_\ell^{(k)}}{T^{(k)} - T^{(k-1)}} - \mu_\ell^{(k)} \right| \le \epsilon$ with high probability. That is,

$$\mathbb{P}\left(N_\ell^{(k)} \ge \left(\mu_\ell^{(k)} - \epsilon_k \right) \left(T^{(k)} - T^{(k-1)} \right) \forall \ell \in \mathcal{I}^{(k)} \right) \ge 1 - \frac{\delta}{mK},\ \epsilon_k \asymp \sqrt{\frac{\tau^{(k)} \log(m^2/\delta)}{T^{(k)} - T^{(k-1)}}}.$$

Summing over all phases until elimination:

$$N_\ell(T_\ell) = \sum_{k:\, \ell \in \mathcal{I}^{(k)}} N_\ell^{(k)} \ge \sum_k \mu_\ell^{(k)} \left(T^{(k)} - T^{(k-1)} \right) - \sum_k \mathrm{err}_k,\ \sum_k \mu_\ell^{(k)} \left(T^{(k)} - T^{(k-1)} \right) = T_\ell \bar{\mu}_\ell^{(\ell)}.$$

The deviation term can be controlled as below:

$$\sum_k \mathrm{err}_k \le \sum_k \sqrt{\tau^{(k)} (T^{(k)} - T^{(k-1)}) \log\left(\frac{m^2}{\delta}\right)} \le \left(\sum_k \tau^{(k)} \right)^{\frac{1}{2}} \left(\sum_k (T^{(k)} - T^{(k-1)}) \log\left(\frac{m^2}{\delta}\right) \right)^{\frac{1}{2}},$$

where the second inequality follows from Cauchy–Schwarz. Thus, $\sum_k \mathrm{err}_k = o(T_\ell)$ when $T^{(k)} \gg \tau^{(k)} \log(\frac{m}{\delta})$. Thus, for sufficiently large T_ℓ, $N_\ell(T_\ell) \ge T_\ell \bar{\mu}_\ell^{(\ell)} - o(T_\ell)$, i.e.,

$$\mathbb{P}(N_\ell(T_\ell) \ge (1 - \epsilon)\, T_\ell \bar{\mu}_\ell^{(\ell)}) \ge 1 - \delta \implies T_\ell \le \frac{N_\ell(T_\ell)}{(1-\epsilon)\bar{\mu}_\ell^{(\ell)}}, T_\ell \approx \frac{N_\ell(T_\ell)}{\bar{\mu}_\ell^{(\ell)}}(1 + o(1))$$

Taking the maximum over $\ell = 1, \ldots, m$, yields the desired result. ■

Corollary 6 *Assume the conditions of Lemma 4 hold. Suppose the empirical time averaged occupancy until elimination for state ℓ, $\bar{\mu}_\ell^{(\ell)}$ which is a weighted average of the stationary distribution of each phase is such that the weights are uniformly bounded, i.e.,*

$$\bar{\mu}_\ell^{(\ell)} = \frac{1}{T_\ell} \sum_{k:\, \ell \in \mathcal{I}^{(k)}} \mu_\ell^{(k)} \left(T^{(k)} - T^{(k-1)} \right) = \sum_{k:\, \ell \in \mathcal{I}^{(k)}} w_k^{(\ell)}\, \mu_\ell^{(k)},\ \textit{where}\ w_k^{(\ell)} = \frac{T^{(k)} - T^{(k-1)}}{T_\ell}, \max_k w_k^{(\ell)} \le C,$$

where $C < 1$. For any $\delta \in (0, 1)$, with probability at least $1 - \delta$, all states are correctly classified after

$$n \gtrsim \left(\max_\ell \tau_{\max}^{(\ell)} \right) \cdot \left(\sum_{\ell=1}^m \frac{1}{\beta_\ell^2} \right) \log\left(\frac{2m}{\delta} \right).$$

Proof From the design problem, Problem 1, we note that $\mu_\ell^{(k)} \propto \frac{1}{\beta_\ell^2}$, i.e., $\mu_\ell^{(k)} \gtrsim \frac{c}{\beta_\ell^2} \quad \forall k$ with $\ell \in \mathcal{I}^{(k)}$. This design problem along with the condition on $w_k^{(\ell)}$ imply that $\bar{\mu}_\ell^{(\ell)} \propto \frac{1}{\beta_\ell^2}$. The early phases of the exploration process cause issues. That is, $\mathcal{I}^{(k)}$ is large, mixing is slow, allocation may be poor. If $T^{(1)} \approx T_\ell$, then $\bar{\mu}_\ell^{(\ell)} \approx \mu_\ell^{(1)}$, so later improvements are irrelevant. The issue is fixed by controlling the phase lengths. Suppose both the following conditions, $T^{(k)} - T^{(k-1)} \gtrsim \tau^{(k)} \log\left(\frac{m}{\delta}\right)$ (A3) and $T^{(k)} - T^{(k-1)} \lesssim \frac{1}{\beta_{\min,k}^2}$ hold. Then no phase dominates. Further, there is a monotone improvement in mixing time, i.e., $\mathcal{I}^{(k+1)} \subset \mathcal{I}^{(k)}$, which implies $\tau^{(k)} \downarrow,\ \mu_\ell^{(k)} \uparrow$. Thus, later phases are better. If

allocations improve, $\mu_\ell^{(k)} \le \mu_\ell^{(k+1)}$, then $\bar{\mu}_\ell^{(\ell)} \ge \min_{k:\, \ell \in \mathcal{I}^{(k)}} \mu_\ell^{(k)}$. In particular, $\bar{\mu}_\ell^{(\ell)} \ge \mu_\ell^{(\text{last phase})}$. Thus, $\bar{\mu}_\ell^{(\ell)} = \frac{c}{\beta_\ell^2}$. Normalizing,

$$\sum_{\ell \in \mathcal{I}^{(k)}} \bar{\mu}_\ell = 1 \quad \Longrightarrow \quad c = \left(\sum_{\ell \in \mathcal{I}^{(k)}} \frac{1}{\beta_\ell^2} \right)^{-1} \Longrightarrow \bar{\mu}_\ell = \frac{\frac{1}{\beta_\ell^2}}{\sum_{j \in \mathcal{I}^{(k)}} \frac{1}{\beta_j^2}} \Longrightarrow \frac{1}{\bar{\mu}_\ell \beta_\ell^2} = \sum_{j \in \mathcal{I}^{(k)}} \frac{1}{\beta_j^2}.$$

Substituting the above into the bound of Theorem 5, the result follows, thus solving Problem 2. ■

We now compare the proposed algorithm with uniform exploration and Bayesian exploration. Uniform exploration induces approximately uniform stationary distribution $\mu_\ell^\pi \approx \frac{1}{m}$, with typically moderate mixing . From Proposition 3, the sample complexity is given by

$$n_{\text{uniform}} \gtrsim \tau_{\text{uniform}} \cdot m \cdot \max_\ell \frac{1}{\beta_\ell^2} \log\left(\frac{2m}{\delta}\right).$$

Bayesian experimental design allocates sampling effort based on uncertainty $\mu_\ell^\pi \propto \frac{1}{\beta_\ell}$, focusing on 'hard" regions (small β_ℓ). Thus,

$$\mu_\ell^\pi \approx \frac{\frac{1}{\beta_\ell}}{\sum_j \frac{1}{\beta_j}} \Longrightarrow \mu_{\min} \approx \left(\sum_j \frac{1}{\beta_j} \right)^{-1} \Longrightarrow n_{\text{Bayes}} \gtrsim \tau_{\text{Bayes}} \cdot \left(\sum_{\ell=1}^m \frac{1}{\beta_\ell} \right) \cdot \max_\ell \frac{1}{\beta_\ell} \log\left(\frac{2m}{\delta}\right).$$

Improvement is achieved when only $k \ll m$ states are difficult, that is, β_ℓ small for k states and large otherwise. Then, $\sum_\ell \frac{1}{\beta_\ell} \approx \frac{k}{\beta_{\min}}$, which implies $n_{\text{Bayes}} \sim \tau_{\text{Bayes}} \cdot \frac{k}{\beta_{\min}^2}$. From Table 1, it follows

Strategy	$\tau_{\max}$	$\mu_{\min}$	Sample complexity
Uniform	moderate	1/m	$\tau \cdot m \cdot \max_\ell \frac{1}{\beta_\ell^2} \cdot \log\left(\frac{2m}{\delta}\right)$
Bayesian	moderate	adaptive	$\tau \cdot \left(\sum_\ell \frac{1}{\beta_\ell}\right) \cdot \max_\ell \frac{1}{\beta_\ell} \cdot \log\left(\frac{2m}{\delta}\right)$
Adaptive Mixing	small	adaptive	$\tau_{\min} \cdot \left(\sum_{\ell=1}^m \frac{1}{\beta_\ell^2}\right) \cdot \log\left(\frac{2m}{\delta}\right)$

Table 1: Comparison of sample complexities of the three different policies.

that the proposed algorithm achieves lower sample complexity when $\beta_l^{(k)} \approx \beta_l$ and the assumptions of Lemma 4 are satisfied. Proposition 2 implies that (A1) is satisfied, while the verification of assumptions (A2) and (A3) is for future investigation.

5. Conclusion and Future Work

We studied the problem of sample-efficient estimation of safety parameters across regions in a given domain. By formulating exploration as the optimization of the spectral gap of the induced Markov chain, we derived a closed-form policy and established high-probability bounds on the sample complexity of the resulting algorithm. The proposed approach demonstrates how adapting the exploration policy to the underlying estimation difficulty and minimizing mixing time can significantly improve efficiency over standard methods. Future work includes extending the framework to continuous state spaces and high-dimensional representations, as well as incorporating task-oriented objectives such as reward maximization and constraint satisfaction. These directions would help unify safety-aware exploration and reinforcement learning, enabling sample-efficient decision-making in more complex environments.

Acknowledgments

This work was supported in part by Swedish Research Council Distinguished Professor Grant 2017-01078, Knut and Alice Wallenberg Foundation Wallenberg Scholar Grant, and the Swedish Strategic Research Foundation FUSS SUCCESS Grant. Aneesh Raghavan thanks Stefan Stojanovic and Alexandre Proutiere for discussion on the problem considered in this paper. The authors thank the anonymous reviewers for their valuable feedback on the initial submission.

References

Alessandro Bottero, Carlos Luis, Julia Vinogradska, Felix Berkenkamp, and Jan R Peters. Information-theoretic safe exploration with gaussian processes. *Advances in Neural Information Processing Systems*, 35:30707–30719, 2022.

Richard C. Bradley. Basic properties of strong mixing conditions. a survey and some open questions. *Probability Surveys*, 2:107–144, 2005.

Kheyreddine Djouzi, Kadda Beghdad-Bey, and Abdenour Amamra. A new adaptive sampling algorithm for big data classification. *Journal of Computational Science*, 61:101653, 2022.

Paul Doukhan. Mixing. In *Mixing: Properties and Examples*, pages 15–23. Springer, 1995.

Jiantao Fan et al. Hoeffding's inequality for general markov chains and its applications. *Journal of Machine Learning Research*, 2021. See JMLR paper (2021) for modern refinements.

Tanner Fiez, Lalit Jain, Kevin G Jamieson, and Lillian Ratliff. Sequential experimental design for transductive linear bandits. *Advances in neural information processing systems*, 32, 2019.

Trygve Olav Fossum, John Ryan, Tapan Mukerji, Jo Eidsvik, Thom Maughan, Martin Ludvigsen, and Kanna Rajan. Compact models for adaptive sampling in marine robotics. *The International Journal of Robotics Research*, 39(1):127–142, 2020.

Adam Foster, Martin Jankowiak, Elias Bingham, Paul Horsfall, Yee Whye Teh, Thomas Rainforth, and Noah Goodman. Variational bayesian optimal experimental design. *Advances in neural information processing systems*, 32, 2019.

Peter W. Glynn and Dirk Ormoneit. Hoeffding's inequality for uniformly ergodic markov chains. *Statistics & Probability Letters*, 56(2):143–146, 2002.

Daolang Huang, Yujia Guo, Luigi Acerbi, and Samuel Kaski. Amortized bayesian experimental design for decision-making. *Advances in Neural Information Processing Systems*, 37:109460–109486, 2024.

Yassir Jedra and Alexandre Proutiere. Sample complexity lower bounds for linear system identification. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pages 2676–2681. IEEE, 2019.

Yassir Jedra, Junghyun Lee, Alexandre Proutiere, and Se-Young Yun. Nearly optimal latent state decoding in block mdps. In *International Conference on Artificial Intelligence and Statistics*, pages 2805–2904. PMLR, 2023.

Pascal Lezaud. Chernoff-type bound for finite markov chains. *The Annals of Applied Probability*, 8(3):849–867, 1998.

Cen-You Li, Olaf Duennbier, Marc Toussaint, Barbara Rakitsch, and Christoph Zimmer. Global safe sequential learning via efficient knowledge transfer. *arXiv preprint arXiv:2402.14402*, 2024.

Sean P. Meyn and Richard L. Tweedie. *Markov Chains and Stochastic Stability*. Cambridge University Press, Cambridge, UK, 2nd edition, 2009.

J. R. Norris. *Markov Chains*. Cambridge University Press, Cambridge, UK, 1997.

Daniel Paulin. Concentration inequalities for markov chains by marton couplings and spectral methods. *Electronic Journal of Probability (and arXiv:1212.2015)*, 2015. See also arXiv:1212.2015.

Martin L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, New York, NY, USA, 1994.

Jens Schreiter, Duy Nguyen-Tuong, Mona Eberts, Bastian Bischoff, Heiner Markert, and Marc Toussaint. Safe exploration for active learning with gaussian processes. In *Joint European conference on machine learning and knowledge discovery in databases*, pages 133–149. Springer, 2015.

Shubhanshu Shekhar, Greg Fields, Mohammad Ghavamzadeh, and Tara Javidi. Adaptive sampling for minimax fair classification. *Advances in Neural Information Processing Systems*, 34:24535–24544, 2021.

Prashant Singh, Joachim van der Herten, Dirk Deschrijver, Ivo Couckuyt, and Tom Dhaene. A sequential sampling strategy for adaptive classification of computationally expensive data. *Structural and Multidisciplinary Optimization*, 55:1425–1438, 2017.

Paul Stankiewicz, Yew T Tan, and Marin Kobilarov. Adaptive sampling with an autonomous underwater vehicle in static marine environments. *Journal of Field Robotics*, 38(4):572–597, 2021.

Yanan Sui, Alkis Gotovos, Joel Burdick, and Andreas Krause. Safe exploration for optimization with gaussian processes. In *International conference on machine learning*, pages 997–1005. PMLR, 2015.

Simon Tong and Daphne Koller. Active learning for parameter estimation in bayesian networks. *Advances in neural information processing systems*, 13, 2000.

Runzhe Wan, Branislav Kveton, and Rui Song. Safe exploration for efficient policy evaluation and comparison. In *International Conference on Machine Learning*, pages 22491–22511. PMLR, 2022.

Patrick Weber, Andrei Kramer, Clemens Dingler, and Nicole Radde. Trajectory-oriented bayesian experiment design versus fisher a-optimal design: an in depth comparison study. *Bioinformatics*, 28(18):i535–i541, 2012.

Bin Zhang and Gaurav S Sukhatme. Adaptive sampling for estimating a scalar field using a robotic boat and a sensor network. In *Proceedings 2007 IEEE International Conference on Robotics and Automation*, pages 3673–3680. IEEE, 2007.