Bandit Quickest Changepoint Detection

Aditya Gopalan
Electrical Communication Engineering
Indian Institute of Science, Bengaluru 560012
aditya@iisc.ac.in

Braghadeesh Lakshminarayanan
Electrical Communication Engineering
Indian Institute of Science, Bengaluru 560012
braghadeesh94@gmail.com

Venkatesh Saligrama
Electrical and Computer Engineering
Boston University, Boston, MA 02215
srv@bu.edu

Abstract

Many industrial and security applications employ a suite of sensors for detecting abrupt changes in temporal behavior patterns. These abrupt changes typically manifest locally, rendering only a small subset of sensors informative. Continuous monitoring of every sensor can be expensive due to resource constraints, and serves as a motivation for the bandit quickest changepoint detection problem, where sensing actions (or sensors) are sequentially chosen, and only measurements corresponding to chosen actions are observed. We derive an information-theoretic lower bound on the detection delay for a general class of finitely parameterized probability distributions. We then propose a computationally efficient online sensing scheme, which seamlessly balances the need for exploration of different sensing options with exploitation of querying informative actions. We derive expected delay bounds for the proposed scheme and show that these bounds match our information-theoretic lower bounds at low false alarm rates, establishing optimality of the proposed method. We then perform a number of experiments on synthetic and real datasets demonstrating the effectiveness of our proposed method.

1 Introduction

We propose a framework for bandit[1] quickest change detection (BQCD). Specifically, we are given a multi-dimensional online data stream, which can be sequentially probed by means of actions belonging to an action set (for instance, only a few among all of the data stream components can be observed at any time, or a linear combination can be acquired). The online data stream can exhibit abrupt *statistical* changes, at any time, and only among a few arbitrary components. Our task is to sequentially probe the data stream by adaptively choosing valid actions based on past bandit[2] observations associated with past actions. Our objective is to detect a change, if it has happened, with minimum detection delay.

Example Scenarios. Surveillance systems [HC11] are equipped with a suite of sensors that can be switched and steered to focus attention on any target or location over a physical landscape (see Fig. 1) to detect abrupt changes at any location. On the other hand, sensor suites are resource limited, and only a limited subset, among all the locations, can be probed at any time. As such,

[1] By 'bandit' we generally mean adaptive sampling with partial information.

[2] This is different from classical quickest change detection [TNB14], where all of the multiple streams are observed at each time, and the only adaptive decision is to declare whether or not a change has happened.

35th Conference on Neural Information Processing Systems (NeurIPS 2021).

we face a fundamental *dilemma*: Focusing attention at any one location can compromise change detection at other locations. Although we describe a scenario in surveillance, the problem of BQCD is general and arises in intrusion detection [Bas99], social networks [Vis+14], disease outbreaks and epidemics [QSC14; YSK15], manufacturing processes [Pur+19a; Din+06], energy limited sensor networks[OGR10; ES10], and vital health monitoring [Vil+17].

In this paper, we derive a fundamental characterization for the delay performance of BQCD, and make explicit, the fundamental tradeoff between delay and false-alarm in the low-false alarm regime. We specifically make the following contributions.

Information-theoretic Lower Bound. We prove a lower bound on the expected detection delay that any BQCD algorithm must suffer at a fixed false alarm rate. The lower bound exhibits a fundamental tradeoff between early stopping (false-alarm) and detection delay (time to detect abrupt change). It offers the key insight that the quickest way to detect a change, at any false-alarm rate, is by playing the 'most informative action', of an 'oracle' who a priori knows the post-change distribution. This suggests that to quickly identify changepoints, we must direct our effort towards rapidly identifying informative actions. On a technical level, we develop a change-of-measure argument for nonstationary, adaptive change detection, that allows for relating the divergences between random trajectories until stopping to the divergence of probability laws under each action for any two problem instances.

ϵ-Greedy Change Detector (ϵ-GCD). We propose ϵ-GCD, which, at a high level, uses a small amount of forced exploration to identify informative actions. The forced exploration allows for rapid convergence towards informative actions, and playing these actions minimizes detection delay. Our ϵ-GCD is based on the generalized max-likelihood/likelihood ratio principle which is utilized to estimate parametric changes. To prove detection delay bounds we draw upon key insights of ϵ-GCD. We first interpret the scheme in terms of competing parallel 'queues', where each queue corresponds to a candidate post-change parameter, collects 'arrivals' which are log-likelihood ratios of observations, and cannot go negative. The true parameter is the queue which enjoys the highest growth rate after a change, and the detection delay is the time required for it to dominate and become the 'longest queue'. The dynamics of the queues can be related to nonstationary random walks with drifts. Using these insights, we prove that the expected detection delay of ϵ-GCD at low false alarm rates mirrors our information theoretic lower bounds thus establishing optimality of our method.

Experiments. We perform numerical experiments of ϵ-GCD on synthetic and real datasets and show that under variations of changepoints, anomalies, and action sets, we realize gains due to adaptive sensing.

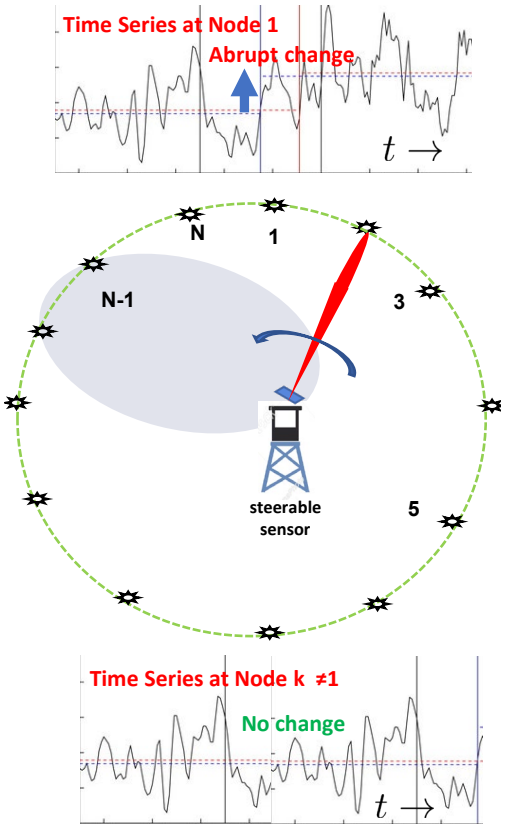

Figure 1: Example BQCD Scenario. There are N spatial-locations that are monitored. The steerable sensor is capable of rapidly switching modes–grey and red beams as shown–and collect aggregated signal from multiple locations at a low signal-to-noise ratio (SNR) or focus attention narrowly on a single location at high SNR. Locations can exhibit abrupt change, which manifests as the depicted time-series (see top). Due to resource constraints, at any time, only a limited set of locations can be observed with high SNR. BQCD aims to learn a policy to adaptively steer and switch modes to different locations and detect abrupt changes if any, quickly.

2 Related Work

Classical quickest change detection, dating back to the pioneering works of Page [Pag54] and Lorden [Lor71], studies the problem of deciding when to stop (akin to declaring a change) while sequentially observing a data stream. In this context, the CUSUM algorithm [Pag54], in

which a change is announced as soon as a likelihood ratio statistic exceeds a threshold, has been shown to enjoy optimal performance.

Our work is motivated by constraints on data collection in multi-stream time-series data, necessitating the design of an adaptive sampling rule in addition to the stopping rule. Similar to our work, Zhang and Mei [ZM20] also propose a sequential sampling method for quickest change detection based on the well-known Shiryaev-Roberts-Pollak scheme (see, e.g., [BN93]). However, their theoretical analysis is rather limited in its scope. In particular, they argue "in principle" that their method will control false alarms for sufficiently large thresholds, and will *eventually* (i.e., after infinite time) choose sensors where changes manifest, without any explicit characterization of detection delays. In contrast, we derive fundamental detection delay under a false alarm rate constraint, thus fully characterizing the BQCD problem. Additionally, their framework is not general enough to handle correlated or structured anomalies, which is a significant aspect of both our theoretical and experimental investigations.

In other related work, Gundersen et al [Gun+21] propose to extend the framework of Bayesian Online Change Detection (BOCD) to incorporate costs in making decisions for real-time data acquisition in multi-fidelity sensing scenarios. Different from this perspective, our approach is frequentist, and does not impose action-specific costs and distributional constraints on the underlying latent parameters or on the changepoints. Furthermore, we derive finite-time performance bounds, which is not a focus of this work.

3 Problem Setup

General Setup. We consider the following discrete-time model for the bandit quickest changepoint detection problem. At each decision round $t \in \mathbb{N} := \{1, 2, \ldots\}$, the parameter θ_t governs the distribution of the observation taken in that round. Let $\nu \in \mathbb{N}$ denote the round at which the change in behavior occurs, so that $\theta_t = \theta_0$ for $t < \nu$ and $\theta_t = \theta^* \in \Theta^{(1)}$ for $t \geq \nu$. At each round t, a learner decides to either (1) stop and output that a change has taken place, denoted by the random variable $U_t = 1$, or (2) continue (denoted by $U_t = 0$) and play an action or sensing decision A_t from an action set $\mathcal{A}$. Upon playing this action the learner obtains an observation X_t, sampled independent of the past, from a probability distribution $\mathbb{P}_{\theta_t}[\cdot|A_t]$, which depends on both the current system parameter θ_t and the current action A_t. The stopping decision and sensing action (U_t and A_t) are assumed to be chosen in a causal manner, i.e., depending on all past information $U_0, A_1, X_1, U_1, \ldots, A_{t-1}, X_{t-1}$, along with potentially independent internal randomness. The learner is aware, a priori, of the pre-change parameter θ_0, the post-change parameter set $\Theta^{(1)}$ and the observation distribution structure, i.e., $\{\mathbb{P}_\theta[\cdot|a]\}_{\theta,a}$ with $a \in \mathcal{A}$ denotes fixed actions. Crucially, the change time ν and the specific post-change parameter θ^* are not known in advance and must be 'learnt' in order to perform well.

The stopping time of the learner is defined to be $\tau = \min\{t \geq 0 : U_t = 1\}$, and the main performance metric that we are interested in is the detection delay, which is the random variable $(\tau - \nu)^+ = \max\{\tau - \nu, 0\}$. Specifically, we are interested in designing sampling and stopping rules so that (a) when no change occurs ($\nu = \infty$), τ should be at least a specified time delay with high probability, which is a measure of false alarm rate, and (b) when a change occurs ($\nu < \infty$), the expected detection delay should be small implying quick detection of change.

Illustrative Example. For illustration, we elaborate on the scenario depicted in Fig. 1. We consider N 'locations' as shown, and the observation at location j at time t follows:

$$Y_j(t) = \theta \mathbb{1}_{[t \geq \nu] \wedge [j \in S]} + W_j(t),\ t \in \mathbb{Z}^+,\ j \in [N],$$

where $\{W_j(t)\}_{j,t}$ is a collection of standard normal random variables and $\mathbb{1}_{[\cdot]}$ denotes the indicator function. This is a temporal extension of the model in [QSC14]. The model with the parameter $\theta \neq 0$ signifies an elevated mean after time $t \geq \nu \in \mathbb{Z}^+$ at locations $S \subseteq [N]$. The learner is agnostic to θ, the time ν as well as which set S among a family of known subsets $\mathcal{S} \subseteq \{0,1\}^N$ exhibits elevated activity. The learner at time t can probe locations by means of sensing actions $A_t \in \mathcal{A} \subset \{0,1\}^N$, and observe a scalar measurement, $X_t \sim P_{\theta_t}(\cdot \mid A_t) \sim \mathcal{N}\left(\theta_t \frac{|A_t \cap S|}{\sqrt{|A_t|}}, 1\right)$, with $\theta_t = 0$ for $t < \nu$ and for $t \geq \nu$, $\theta_t = \theta \in \Theta^{(1)} = \mathbb{R} \setminus \{0\}$. The normalization factor $\sqrt{|A_t|}$ expresses the fact that the same SNR is maintained across different probes A_t (see [QSC14]).

General (Structured) Scenarios. Depending on the application, locations maybe endowed with a neighborhood network structure, and change maybe observable in a K-neighborhood of the target location, and in such cases, $\mathcal{S}$ are constrained to be K-neighborly sets. Similarly, to allow for diverse probing modes, we could allow the action set to include actions that aggregate signal across multiple locations.

Notation. For a post-change parameter $\theta \in \Theta^{(1)}$, we use $a^*(\theta)$ to denote the most informative action for θ vs. $\theta^{(0)}$, i.e., $a^*(\theta) \in \operatorname{argmax}_{a\in\mathcal{A}} D\left(\mathbb{P}_\theta\left[\cdot|a\right]||\mathbb{P}_{\theta_0}\left[\cdot|a\right]\right)$, where $D\left(\cdot||\cdot\right)$ stands for the Kullback-Leibler divergence. To lighten notation, in the sequel we often abbreviate the distribution $\mathbb{P}_\theta\left[\cdot|a\right]$ to $\theta(a)$, and thus $D\left(\mathbb{P}_\theta\left[\cdot|a\right]||\mathbb{P}_{\theta_0}\left[\cdot|a\right]\right)$ to $D\left(\theta(a)||\theta_0(a)\right)$. The simultaneous subscript-superscript notation z_i^j is often used to represent the sequence $(z_i, z_{i+1}, \ldots, z_j)$. For any $\theta \in \Theta^{(1)}$ and $n \in \mathbb{N}$, we use $\mathbb{P}^{(n,\theta)}$ to denote the probability measure on the process $U_0, A_1, X_1, U_1, A_2, X_2, \ldots$ induced by the learner's decisions, when the pre-change parameter θ_0 changes to θ at time n. We also denote by $\mathbb{P}^{(\infty)}$ the probability measure as above under the no-change situation ($\nu = \infty$).

We define the filtration $(\mathcal{F}_t)_{t\geq 0}$ to be the σ-algebra generated by all the past random variables up to time t, i.e., $\mathcal{F}_t = \sigma(\{(A_s, X_s), 0 \leq s \leq t)\})$. An (admissible) change detector is an algorithm for which the stopping decision $U_t \in \{0, 1\}$ and action $A_t \in \mathcal{A}$ are $\mathcal{F}_{t-1}$-measurable at all rounds t.

Objective Specification. Our BQCD framework involves a false alarm rate constraint specified by a tuple (m, α), where $m \in \mathbb{Z}^+$ denotes an a priori fixed time such that the probability of an admissible change detector stopping before time m under the null hypothesis (no change) is at most $\alpha \in (0, 1)$, namely, $\mathbb{P}^{(\infty)}\left[\tau < m\right] \leq \alpha$. This kind of false alarm duration specification for the stopping time ($\geq m$ with probability $\geq 1 - \alpha$) has been commonly employed in the literature on classical (non-adaptive) change detection (see e.g., [Lai98]), to ensure that the comparative performance rules out early triggering algorithms. Let $\mathcal{D}_{m,\alpha}$ be the class of admissible change detectors with the (m, α) property stated above. As such, our goal is to find detectors in $\mathcal{D}_{m,\alpha}$ that minimize the detection delay when a change indeed occurs.

Informal Statement of Our Results. Our main result is a sharp "optimal characterization" for expected detection delay in the limit of vanishing false alarms ($\alpha \to 0$, $m \to \infty$). The result, stated below, involves the new ϵ-Greedy-change-detector (ϵ-GCD) algorithm in which, in each round t, we choose an action uniformly at random from the set $\mathcal{A}$ with probability ϵ, or choose an action that is 'most informative' for an estimated post-change parameter with probability $1 - \epsilon$. **Objective Specification.** Our BQCD framework involves a false alarm rate constraint specified by a tuple (m, α), where $m \in \mathbb{Z}^+$ denotes an a priori fixed time such that the probability of an admissible change detector stopping before time m under the null hypothesis (no change) is at most $\alpha \in (0, 1)$, namely, $\mathbb{P}^{(\infty)}\left[\tau < m\right] \leq \alpha$. This kind of

Theorem 1 (informal)**.** *Consider the limiting situation as $\alpha \to 0$ and $m = \omega\left(\log \frac{1}{\alpha}\right) \to \infty$. Then, any admissible change detector in $\mathcal{D}_{m,\alpha}$ must suffer*

$$\liminf_{\alpha\to 0} \frac{\mathbb{E}^{(\nu,\theta^*)}[(\tau - \nu)^+]}{\log \frac{1}{\alpha}} \geq \frac{1}{\max_a D(\theta^*(a)||\theta^{(0)}(a))}.$$

On the other hand, the ϵ-GCD algorithm, suitably tuned to be in $\mathcal{D}_{m,\alpha}$, achieves

$$\limsup_{\alpha\to 0} \frac{\mathbb{E}^{(\nu,\theta^*)}[(\tau - \nu)^+]}{\log \frac{1}{\alpha}} \leq \frac{8}{(1-\epsilon)} \frac{1}{\max_a D(\theta^*(a)||\theta^{(0)}(a))}.$$

4 Fundamental lower bounds on bandit change detection performance

Before embarking upon the design of bandit changepoint algorithms, it is worth understanding what the limits of bandit change detection performance are, due to the stochastic and partial nature of the problem's information structure. We expose in this section a universal lower bound on the detection delay of *any* bandit changepoint algorithm with a false alarm rate above a given level.

Theorem 2. *Let $0 < \alpha \leq \frac{1}{10}$ and $m \geq 1$. For any bandit changepoint algorithm satisfying $\mathbb{P}^{(\infty)}\left[\tau < m\right] \leq \alpha$, we have $\mathbb{E}^{(1,\theta^*)}\left[\tau\right] \geq \min\left\{\frac{\frac{1}{20}\log\frac{1}{\alpha}}{\max_{a\in\mathcal{A}} D\left(\theta^*(a)||\theta^{(0)}(a)\right)}, \frac{m}{2}\right\} \quad \forall \theta^* \in \Theta^{(1)}$.*

The result implies that in the 'small' false alarm rate (α) regime with 'large' $m = \omega(\log(1/\alpha))$, e.g., $m = \log^2(1/\alpha)$, we have that any algorithm meeting the false alarm rate property $\mathbb{P}^{(\infty)}\left[\tau \geq m\right] \geq$

$1-\alpha$ must suffer a detection delay at least $\Omega\left(\frac{\log\frac{1}{\alpha}}{\max_{a\in\mathcal{A}} D\left(\theta^*(a)||\theta^{(0)}(a)\right)}\right)$ when a change occurs at time 1 (a similar, 'anytime' lower bound holds for detection delay at $\nu \in \mathbb{N}$, see appendix).

False alarm-detection Delay Tradeoff. Theorem 2 shows that a basic tradeoff exists between the false alarm rate or early stopping rate in case of no change, on one hand, and the detection delay after a true change occurs, on the other. Specifically, it is impossible to stop 'too early', i.e., before time $\Omega\left(\frac{\log\frac{1}{\alpha}}{\max_{a\in\mathcal{A}} D\left(\theta^*(a)||\theta^{(0)}(a)\right)}\right)$, after a true change if one wishes to stop 'too late' under no change, i.e., $\mathbb{P}^{(\infty)}\left[\tau \geq m\right] \geq 1-\alpha$. We also note that when there is no adaptive sampling of actions (i.e., $|\mathcal{A}| = 1$), then the lower bound reduces to the form of a a standard lower bound for the classical changepoint detection problem [Lai98].

Information Structure. This is perhaps the most valuable implication of the lower bound. The quantity $D\left(\theta^*(a)||\theta^{(0)}(a)\right)$, in the denominator, can be interpreted as the amount of information that playing an action a provides (on average) in order to detect a change from θ_0 to θ^*. Consequently, the quickest way to detect a change is by playing the 'most informative action', $\operatorname{argmax}_{a\in\mathcal{A}} D\left(\theta^*(a)||\theta^{(0)}(a)\right)$. This can be viewed as the sampling strategy of an 'oracle' who knows the post-change parameter θ^* in advance. Unfortunately, this information is not known a priori for a causal algorithm. However, we will see that this can in fact be learnt along the trajectory and the lower bound can be attained order-wise by a suitable learning algorithm that we propose.

Proof Sketch for Theorem 2. The main idea is a measure change argument, adapted to our non-stationary change point setting, from literature on sample complexity bounds for bandit best arm identification [GK16]. On one hand, if the algorithm in consideration is very 'lazy' to begin with, i.e., $\mathbb{P}^{(1,\theta^*)}\left[\tau \geq m\right]$ is at least a constant, say $1/2$, then we immediately get $\mathbb{E}^{(1,\theta^*)}\left[\tau\right] \geq \frac{1}{2}\cdot m$. On the other hand, if the algorithm is 'active', i.e., $\mathbb{P}^{(1,\theta^*)}\left[\tau < m\right] \geq 1/2$, then the KL divergence between the laws of the indicator random variable $\mathbf{1}\left\{\tau < m\right\}$ in the instances $(1,\theta^*)$ and $(\infty,\cdot)$ is 'large' (at least a constant times $\log(1/\alpha)$) owing to the hypothesis that $\mathbb{P}^{(\infty)}\left[\tau < m\right]$ is 'small' (at most α). But by the data processing inequality, this divergence is bounded above by the divergence between the distributions of the entire trajectory $U_0, A_1, X_1, U_1, A_2, X_2, \ldots$ under the two instances, which can be seen to be equivalent to the quantity $\sum_{a\in\mathcal{A}} D\left(\theta^*(a)||\theta(a)\right)\mathbb{E}^{(1,\theta^*)}\left[N_\tau(a)\right]$, and which is further bounded above by $\mathbb{E}^{(1,\theta^*)}\left[\tau\right]\max_{a\in\mathcal{A}} D\left(\theta^*(a)||\theta(a)\right)$. See appendix for a complete proof.

5 ϵ-GCD Algorithm

We describe the ϵ-GCD adaptive sensing algorithm for the bandit quickest change detection problem. The lower bound in Section 4 suggests that it is beneficial to infer the target post-change parameter, so that playing the most informative action for it can yield the best possible detection delay performance. This is the key principle underlying the design of the ϵ-GCD algorithm (Algorithm 1).

At a high level, ϵ-GCD uses a small amount of forced exploration along with 'greedy' exploitation to play sensing actions. Specifically, it computes, at each round t, a maximum likelihood estimate (MLE) of the post-change parameter based on the generalized likelihood ratio test (GLRT) principle. It then plays either a randomly chosen action, if the current slot is an exploration slot, or a 'greedy', i.e., most informative, action for the estimated post-change parameter, if it is an exploitation slot.

The MLE of the post-change parameter, in round t, admits an interpretation as the longest 'queue' corresponding to some parameter in $\Theta^{(1)}$. To see this, notice that the MLE for the pair (ν,θ^*), given all previous data in exploration rounds can be written as $\operatorname{argmax}_{\theta\in\Theta^{(1)},1\leq v\leq t}\prod_{\ell=1}^{v-1} f_{\theta_0}(X_\ell|A_\ell)^{E_\ell}\prod_{\ell=v}^{t-1} f_\theta(X_\ell|A_\ell)^{E_\ell} = \operatorname{argmax}_{\theta\in\Theta^{(1)},1\leq v\leq t}\sum_{\ell=v}^{t-1} E_\ell \log\frac{f_\theta(X_\ell|A_\ell)}{f_{\theta_0}(X_\ell|A_\ell)}$, with $f_\theta(\cdot|a)$ taken to be the probability density or mass function of the distribution $\mathbb{P}_\theta(\cdot|A_\ell)$ (assuming one exists). Observe now that for each candidate post-change parameter θ, the inner maximum over v, $Q_t^{(1)}(\theta) := \operatorname{argmax}_{1\leq v\leq t}\sum_{\ell=v}^{t-1} E_\ell \log\frac{f_\theta(X_\ell|A_\ell)}{f_{\theta_0}(X_\ell|A_\ell)}$, evolves as a 'queue'[3] with arriving work $E_\ell \log\frac{f_\theta(X_\ell|A_\ell)}{f_{\theta_0}(X_\ell|A_\ell)}$ at time slot ℓ; in other words, $Q_{t+1}^{(1)}(\theta) = \left(Q_t^{(1)}(\theta) + E_t \log\frac{f_\theta(X_t|A_t)}{f_{\theta_0}(X_t|A_t)}\right)^+$.

[3]This is also known as the Lindley recursion equation in queueing theory.

We define the algorithm using general 'processing' functions g_θ in place of the log-likelihood ratios $\log \frac{f_\theta}{f_{\theta_0}}$ above. Broadly speaking, the functions g should ideally be chosen with the hope that (a) $g_\theta(X_\ell, A_\ell)$ is negative in expectation before the change time ($\ell < \nu$), and (b) $g_\theta(X_\ell, A_\ell)$ is large and positive in expectation for $\theta = \theta^*$, the true change parameter, after the change ($\ell \geq \nu$).

The stopping rule that ϵ-GCD uses is based on the generalized likelihood ratio (GLR)-type statistic. It is the largest of an ensemble of evolving exploitation-data queues $Q_t^{(2)}(\theta)$, which mirrors the definition of $Q_t^{(1)}(\theta)$ but with $1 - E_t$ instead of E_t.

5.1 Theoretical guarantees for the ϵ-GCD algorithm

In this section, we present and discuss theoretical guarantees on the false alarm rate and detection delay performance of the ϵ-GCD algorithm of Section 5.

Algorithm 1 ϵ-GCD

Input: $\epsilon \in [0,1]$, θ_0, $\Theta^{(1)}$, $\beta \geq 1$, π, Observation function $g_\theta(x,a)\ \forall(\theta,x,a) \in \Theta^{(1)} \times \mathcal{X} \times \mathcal{A}$.
Init: $Q_1^{(1)}(\theta) \leftarrow 0$, $Q_1^{(2)}(\theta) \leftarrow 0\ \forall\theta \in \Theta^{(1)}$ {CUSUM statistics for exploration and exploitation per θ}
for round $t = 1, 2, 3, \ldots$ **do**
  if $\max_{\theta\in\Theta^{(1)}} Q_t^{(2)}(\theta) \geq \log\beta$ **then**
    break {Stop sampling and exit}
  end if
  Sample $E_t \sim \text{Ber}(\epsilon)$ independently
  if $E_t == 1$ **then**
    Play action $A_t \sim \pi$ independently {Explore}
    Get observation X_t
    Set $\forall\theta \in \Theta^{(1)} : Q_{t+1}^{(1)}(\theta) \leftarrow \left(Q_t^{(1)}(\theta) + g_\theta(X_t, A_t)\right)^+$ {Update exploration CUSUM statistics}
  else
    Compute $\hat{\theta}_t = \text{argmax}_{\theta\in\Theta^{(1)}} Q_t^{(1)}(\theta)$ {Most likely post-change distribution based on exploration data}
    Play action $A_t = \text{argmax}_{a\in\mathcal{A}} D\left(\theta^{(0)}(a)||\hat{\theta}_t(a)\right)$
    Get observation X_t
    Set $\forall\theta \in \Theta^{(1)} : Q_{t+1}^{(2)}(\theta) \leftarrow \left(Q_t^{(2)}(\theta) + g_\theta(X_t, A_t)\right)^+$ {Update exploitation CUSUM statistics}
  end if
end for

We make the following assumptions on the parameter space and observation distributions in order to derive performance bounds for the algorithm.

Assumption 1. *The post-change parameter set $\Theta^{(1)}$ is finite, i.e., $|\Theta^{(1)}| < \infty$.*

This assumption is made primarily for ease of analysis, whereas the algorithm is defined even for arbitrary parameter spaces. Specifically, it allows for easy control of the fluctuations of an ensemble of (drifting) random walks via a union bound over the parameter set. While we believe that this can be relaxed to handle general parameter spaces via appropriate netting or chaining arguments, this chiefly technical task is left to future investigation.

Assumption 2. *Every post-change parameter is detectable by some action, i.e., $\forall\theta \in \Theta^{(1)}\ \exists a \in \mathcal{A} : D\left(\theta(a)||\theta^{(0)}(a)\right) > 0$. Moreover, any two post-change parameters are separable by some action, i.e., $\forall\theta, \theta' \in \Theta^{(1)}\ \exists a \in \mathcal{A} : D\left(\theta(a)||\theta'(a)\right) > 0$.*

This is a basic identifiability requirement of the setting, without which some parameter changes could be completely undetectable. Put differently, one can only hope to detect changes that the sensing set can tease apart.

Assumption 3 (Bounded marginal KL divergences)**.** *There is a constant $D_{\max}$ satisfying $D\left(\theta_1(a)||\theta_2(a)\right) \leq D_{\max}$ for each $a \in \mathcal{A}$, $\theta_1, \theta_2, \in \Theta^{(1)}$.*

This assumption is easily met, for instance, if all log likelihood ratios are bounded, or if the observations are modelled as Gaussians. In this case log likelihood ratios are linear functions of the observation.

Assumption 4. *Every observation probability distribution $\mathbb{P}_\theta[\cdot|a]$, for $a \in \mathcal{A}$ and $\theta \in \Theta^{(1)} \cup \{\theta_0\}$, has either a density[4] or a mass function, denoted by $f_\theta(\cdot|a)$. Moreover, there exists $r > 0$ such that for any $\theta, \theta', \theta'' \in \Theta^{(1)} \cup \{\theta_0\}$, and $a \in \mathcal{A}$, the log likelihood ratio $\frac{f_{\theta'}(X|a)}{f_{\theta''}(X|a)}$ is r-subgaussian [5] under $X \sim \mathbb{P}_\theta[\cdot|a]$.*

[4] with respect to a standard reference measure, e.g., Lebesgue measure

[5] A random variable X is r-subgaussian under $X \sim \mathbb{P}$ if $\forall\lambda \in \mathbb{R}$: $\mathbb{E}[e^{\lambda(X-\mathbb{E}[X])}] \leq e^{\lambda^2 r/2}$.

This assumption is common in the statistical inference literature; we use it to be able to control the fluctuations of the exploration and exploitation CUSUM statistics in the algorithm via standard subgaussian concentration tools. A concrete example of a setting that meets Assumptions 1-4 is the linear observation model described in Section 3 and Fig. 1. We are now in a position to state our key theoretical result.

Theorem 3 (False alarm and detection delay for general change point)**.** *Under assumptions 1-4, the following conclusions hold for ϵ-GCD (Algorithm 1) run with the log-likelihood observation function $g_\theta(x,a) \equiv \log \frac{f_\theta(x|a)}{f_{\theta_0}(x|a)}$.*

1. (Time to false alarm) Let $\alpha \in (0,1)$ and $m \in \mathbb{N}$. If the stopping threshold β is set as $\beta \geq \frac{m|\Theta^{(1)}|}{\alpha}$, then the stopping time satisfies $\mathbb{P}^{(\infty)}[\tau < m] \leq \alpha$.
2. (Detection delay) For a change from θ_0 to $\theta^* \in \Theta^{(1)}$ occurring at time $\nu \in \mathbb{N}$,

$$\mathbb{E}^{(\nu,\theta^*)}\left[(\tau-\nu)^+\right] \leq \frac{8\log\beta}{(1-\epsilon)\max_a D\left(\theta^*(a)||\theta^{(0)}(a)\right)} + O\left(poly\left(\frac{1}{\epsilon}\right), \mathbb{E}^{(\infty)}\left[Q_\nu^{(1)}\right], \mathcal{P}\right), \tag{1}$$

provided $\pi(a^*_{\theta^*}) > 0$, with $Q_\nu^{(1)} := \left(Q_\nu^{(1)}(\theta)\right)_{\theta\in\Theta^{(1)}}$ denoting the explore queue statistics for all parameters and $\mathcal{P} \equiv (\mathbb{P}_\theta[\cdot|a])_{\theta\in\{\theta_0\}\cup\Theta^{(1)}}$ all the observation distributions.

Remark. We note that β is an internal, algorithmic parameter, whose setting is spelt out in order that the (m,α) false alarm constraint can be met. Thus, the first part of Theorem 3 is a 'correctness' statement[6]

Interpretation of the Detection Delay Bound. The first term in the detection delay bound (1) can be interpreted as the time for an *oracle* fixed-action strategy (e.g., CUSUM), that always plays the most informative action for θ^*, to stop. The second term, on the other hand, is a bound on the time taken by the algorithm to *learn* to play the most informative action for θ^* (details follow in proof sketch). We omit the precise dependence of the second term on the problem structure here for brevity, but detail it explicitly in the appendix. Specifically, for small forced exploration rates ϵ, the second term depends on the information geometry of the problem approximately as $\frac{1}{\epsilon^4}\sum_{\theta\in\Theta^{(1)}:a^*_\theta\neq a^*_{\theta^*}} \frac{1}{\Delta_\theta^4}$. Here, for each candidate post-change parameter $\theta \in \Theta^{(1)}$, its 'gap' Δ_θ is a measure of how difficult it is for the algorithm to eliminate θ as an estimate for the true post-change parameter θ^* during this parameter-learning phase. We formally define it as $\Delta_\theta := \bar{D}_{\theta^*,\theta_0} - \frac{1}{2}(\bar{D}_{\theta^*,\theta_0} + \left(\bar{D}_{\theta^*,\theta_0} - \bar{D}_{\theta^*,\theta}\right)^+)$, where $\bar{D}_{\theta^*,\theta} := \sum_{a\in\mathcal{A}} \pi(a) D\left(\theta^*(a)||\theta(a)\right)$ is the average information divergence between parameters θ^*, θ offered by playing from the exploration distribution π. With finer analysis, the dependence on gaps and ϵ could be improved. We omit it for simplicity.

We now turn to the (additive & linear; see the appendix) dependence on $\mathbb{E}^{(\infty)}\left[Q_\nu^{(1)}\right]$ of the second term. The insight as to why this arises is as follows. The 'learning' phase for θ^*, after a change takes place at time ν, can be seen as a 'race' between competing log-likelihood queues of all the parameters vying to become the MLE $\hat{\theta}_t$. At the change time, each non-ground truth queue $Q_\nu^{(1)}(\theta)$ is, on average, at level $\mathbb{E}^{(\infty)}\left[Q_\nu^{(1)}(\theta)\right]$, representing its initial 'advantage' over the ground-truth queue $Q_\nu^{(1)}(\theta^*)$ which in the worst case could be at level 0. Thus, the difference $\mathbb{E}^{(\infty)}\left[Q_\nu^{(1)}(\theta)\right] - Q_\nu^{(1)}(\theta^*) \leq Q_\nu^{(1)}(\theta)$ is the extra amount of 'time work' that the θ^* queue must do to overcome the θ queue. A special case of the result is for $\nu = 1$ in which case all queues start out at 0. For general ν, one can establish that $\mathbb{E}^{(\infty)}\left[Q_\nu^{(1)}(\theta)\right]$ must be finite, by noticing that each queue is non-negative and accumulates arrivals with negative mean before the change, and applying standard queueing stability arguments. This also indicates that the detection delay bound is roughly invariant given the location of the change point ν.

Optimality at low false-alarms. The result implies that in the 'small' time to false alarm regime, i.e., $\alpha \to 0$, $m = \omega(\log(1/\alpha))$ and $\log(m) = o(\log(\alpha))$, the detection delay when β is set as

[6]This is, at a high level, akin to specifying a PAC algorithm's correctness requirement and measuring its sample complexity on the other hand, i.e., insist that its returned answer be ϵ-correct with probability at least $1-\delta$, and then analyze how long it takes to learn and return an answer. The former corresponds to the false alarm constraint and the latter to measuring detection delay here.

above is dominated by the first term: $\mathbb{E}^{(\nu,\theta^*)}[\tau] = O\left(\frac{\log\beta}{(1-\epsilon)\max_a D\left(\theta^*(a)||\theta^{(0)}(a)\right)}\right)$. This matches, order-wise, the universal lower bound of Theorem 2 up to the algorithm-dependent multiplicative factor $1/(1-\epsilon)$, which may be interpreted as a 'penalty' for the forced exploration mechanism.

Benefits of Adaptivity. Let us consider a concrete setting to elucidate the detection delay bound. In this example, there are d sensing actions, represented by the canonical basis vectors in $\mathbb{R}^d$. The pre-change parameter is $0 \in \mathbb{R}^d$, and there are d candidate post-change parameters, each of which is a canonical basis vector e_i, $i \in [d]$, multiplied by a constant δ. The observation from applying action a at system parameter θ is Gaussian distributed with mean $\langle a, \theta\rangle$ and variance 1. Thus, the aim is to detect a (sparse) change of magnitude δ in one of the d coordinates in $\theta_0 = 0$, when at any time only a single coordinate can be noisily sensed. Suppose $\theta^* = \delta e_1$ without loss of generality, for which the most informative action (in fact, the only informative one) is $a^*_{\theta^*} = e_1$. The first term in the detection delay given by Theorem 3 is then $O\left(\frac{\log(1/\alpha)}{(1-\epsilon)\delta^2}\right)$. This is a factor of d smaller than $\Omega\left(d \cdot \frac{\log(1/\alpha)}{(1-\epsilon)\delta^2}\right)$ that a standard CUSUM rule with non-adaptive uniform sampling over coordinates would achieve – this is seen by applying the lower bound (Theorem 2) to the case of a single (trivial action) where the divergence in the denominator reduces to the *average* divergence across all coordinates: $O(\delta^2/d)$. For estimating the second term, we calculate that for each $\theta = \delta e_i$ with $i \neq 1$, $\Delta_\theta = \bar{D}_{\theta^*,\theta_0} - \frac{(\bar{D}_{\theta^*,\theta_0} + \left(\bar{D}_{\theta^*,\theta_0} - \bar{D}_{\theta^*,\theta}\right)^+)}{2} = \frac{\delta^2}{2d} - \frac{\frac{\delta^2}{2d} - \left(\frac{\delta^2}{2d} - \frac{\delta^2}{d}\right)^+}{2} = \frac{\delta^2}{2d} - \frac{\frac{\delta^2}{2d} - 0}{2} = \frac{\delta^2}{4d}$. Thus, the second term representing the time to learn the optimal sensing action scales as $O\left(\frac{(d-1)d^4}{\epsilon^4\delta^8}\right) = O\left(\frac{d^5}{\epsilon^4\delta^8}\right)$. Although we have not optimized dependence on d, ϵ and δ, the overall bound still gives an idea of how the active sensing problem gets harder as the dimension d grows or as the minimum change amount δ changes, making it akin to finding a 'needle in a haystack'.

Sketch of the Proof of Theorem 3. This section lays down the key arguments involved in the proof of Theorem 3 for bounding the false alarm time and detection delay of the ϵ-GCD algorithm.

1. Bounding the probability of early stopping under no change. The stopping time τ for ϵ-GCD, by line 4 in Algorithm 1, is equivalent to the first instant t when the worst-case (over $\Theta^{(1)}$) CUSUM statistic computed over exploitation data, i.e., $\max_{\theta\in\Theta^{(1)}} Q_t^{(2)}(\theta) \equiv \max_{\theta\in\Theta^{(1)}} \max_{1\le s\le t} \prod_{\ell=s}^{t-1} \left(\frac{f_\theta(X_\ell|A_\ell)}{f_{\theta_0}(X_\ell|A_\ell)}\right)^{(1-E_\ell)}$ exceeds the level $\beta \geq 1$. Under no change, each observation X_ℓ is distributed as $\mathbb{P}_{\theta_0}[\cdot|A_\ell]$ given A_ℓ, and the product $\prod_{\ell=s}^{t-1}\left(\frac{f_\theta(X_\ell|A_\ell)}{f_{\theta_0}(X_\ell|A_\ell)}\right)^{(1-E_\ell)}$, over $t = s, s+1, s+2, \ldots$, behaves as a (standard) likelihood ratio martingale with unit expectation under the appropriate filtration. Hence, the chance that the largest among this ensemble of martingales, one for each $\theta \in \Theta^{(1)}$ and starting time $s \in [m]$, rises above β before time m is bounded using a union bound and Doob's inequality for each individual martingale, yielding the probability bound $\frac{m\left|\Theta^{(1)}\right|}{\beta}$.

2. Control of the detection delay. Suppose that the change takes place at time $\nu = 1$ to the post-change parameter $\theta^* \in \Theta^{(1)}$. The proof strategy is to show, with high probability, that τ is no more than $t_1 + t_2$, where

- t_1 is an upper bound for the time taken for the plug-in estimate $\hat{\theta}_t$, of the post-change parameter, to 'settle' to θ^*. In other words, we show that after time t_1 it is very unlikely that an action other than $a^*_{\theta^*} = \text{argmax}_{a\in\mathcal{A}} D\left(\theta^{(0)}(a)||\theta^*(a)\right)$ is played in every exploitation round.
- t_2 is an upper bound for the time taken for the worst-case CUSUM statistic $\max_{\theta\in\Theta^{(1)}} Q_t^{(2)}(\theta)$ to grow to the level β *assuming that the optimal (i.e., most informative) action $a^*_{\theta^*}$ is always played at all exploitation rounds*.

For clarity of exposition, we will assume that we are in the setting of *linear measurements* in $\mathbb{R}^d$ with *additive standard Gaussian noise* – this makes KL divergences easy to interpret as Euclidean distances – and that the changepoint is at $\nu = 1$.

Step 1: Finding t_1 (time to learn θ^*). We observe that the MLE $\hat{\theta}_t$ at time t can be written as the parameter θ associated with the largest stochastic process $\max_{v=1}^{t-1} \log J_{v,t-1}(\theta)$, one for each θ: $\hat{\theta}_t = \text{argmax}_{\theta\in\Theta^{(1)}} \max_{v=1}^{t-1} \log J_{v,t-1}(\theta)$, where $J_{v,t-1}(\theta) := \prod_{\ell=v}^{t-1}\left(\frac{f_\theta(X_\ell|A_\ell)}{f_{\theta_0}(X_\ell|A_\ell)}\right)^{E_\ell}$.

Under the post-change distribution $\mathbb{P}^{(1,\theta^*)}$, $\log J_{v,t-1}(\theta)$ can be seen to evolve (as a function of t) as a random walk with drift $\epsilon(\|\theta^* - \theta\|_H^2 - \|\theta\|_H^2)$. Here, $\|\cdot\|_H$ is the usual matrix-weighted Euclidean norm in $\mathbb{R}^d$, governed by how much different directions are explored in expectation[7] by the exploration distribution π: $H = \sum_{a \in \mathcal{A}} \pi(a) a a^T$. Note that $\|\theta^* - \theta\|_H^2$ here is the KL divergence between the distributions of the observation that results when an action sampled from π is played, under parameters θ^* and θ.

The preceding discussion implies that the drift is the largest (and positive) for the random walk corresponding to the true post-change parameter $\theta = \theta^*$. The remainder of this part of the proof uses martingale concentration tools (subgaussian Chernoff and maximal Hoeffding bounds) to find the time t_1 at which the fastest-growing random walk, $\log J_{v,t}(\theta^*)$ dominates all other 'competing' random walks $\log J_{v,t}(\theta)$, $\theta \neq \theta^*$.

Step 2: Finding t_2 (time to stop under optimal action plays). In a manner similar to that of Step 1, we observe that the logarithm of the log-likelihood ratio for θ^* w.r.t. θ_0 computed over only *exploitation* rounds, i.e., $\log \prod_{\ell=v}^{t-1} \left(\frac{f_\theta(X_\ell|A_\ell)}{f_{\theta_0}(X_\ell|A_\ell)} \right)^{1-E_\ell}$, evolves as a random walk with drift rate $(1-\epsilon)\|\theta^*\|_H$. A Chernoff bound can then be used to control the probability of the 'bad' event that this growing random walk has *not* crossed the level $\log \beta$ in a certain time duration t_2 (t_2 does not appear explicitly in the main derivation).

6 Experiments

Our goal in this section is to illustrate various aspects of our theory through experiments on synthetic datasets and explore performance of the proposed ϵ-GCD algorithm on a setting based on a real-world dataset. All experiments were performed on a laptop with Intel Core i5 CPU and 8GB of RAM, and take under an hour to execute. The closely related prior works on this topic [ZM20; Gun+21] are more specific and are not compatible with many of our structured scenarios (see Sec. 2. As alternatives, we propose several baselines and compare against our method to benchmark performance.

Size	Oracle	ϵ-GCD+	ϵ-GCD	URS
10	30 ± 5	98 ± 62	112 ± 74	306 ± 76
15	31 ± 5	129 ± 95	158 ± 119	460 ± 115
20	31 ± 5	163 ± 128	196 ± 156	616 ± 153
25	31 ± 5	191 ± 154	253 ± 216	764 ± 194

Table 1: Observed mean and standard deviation for the simulated stopped time for varying graph-sizes with pointy action set $\mathcal{A}_1$, for change occurring at $\nu = 40$.

Uniform Random Sampling (URS). At every round we sample each action in $\mathcal{A}$ uniformly at random. Our goal with URS is to calibrate adaptation gains.
ϵ-GCD (Ours). This is our proposed scheme. A point to be noted here (as seen from Algorithm 1 is that the data used for making a stopping decision ("exploitation") and that for estimating the changed distribution ("exploration") are non-overlapping, and as such leads to inefficient data usage.
ϵ-GCD+. To reduce the aforementioned data inefficiency, we consider the variant, the parameter estimation part also utilizes both data collected during exploration rounds and the data collected for the exploitation rounds.
Oracle. Here we consider an omniscient strategy in that the algorithm is aware of both the pre-change and post-change distributions ahead of time, and consequently does not require exploration to estimate the post-change parameter. On the other hand the method has no knowledge of stopping time, and thus the problem reduces to the conventional quickest change detection. The Oracle characterizes the gold-standard in that no adaptation is required, and it gives the absolute smallest detection delay for any BQCD problem.

Synthetic Experiments. We conduct synthetic experiments for the model introduced in Section 3 under the bold item titled "illustrative example." We explore how average detection delay varies under controlled variation of various parameters such as changepoint location, dimensionality and structure of the ambient space, and the type of action sets. Our objective is twofold: (a) Illustrate gains due to adaptivity of proposed ϵ-GCD over the non-adaptive method, where actions are chosen uniformly at random; (b) Demonstrate "near" optimality by baselining against Oracle.

[7] This matrix also arises commonly in linear design of experiments as the Kiefer-Wolfowitz matrix [LS20].

We report results for the case when the ambient space is a line graph (see Illustrative example in Sec. 3). Nodes are interpreted as physical locations, and take values in $[N]$. Nodes j, k are connected if $|j - k| = 1$. Each node $n \in [N]$ offers a Gaussian-distributed observation depending on the changepoint $\nu \in \mathbb{N}$.

Isolated and Structured Anomalies. The vector of change parameters $\theta = [\theta_n]$ are of two types: (a) Isolated singleton change, namely, $\theta_n \in \{0, 1\}$ and $\sum_{n\in[N]} \theta_n = 1$; (b) Structural changes, i.e., $\text{Supp}(\theta) = \{n \in [N] : \theta_n \neq 0\}$ is k-connected set with $\theta \in \{0, 1\}^N$.

Diffuse and 'Pointy' Action sets. In a parallel fashion we consider two types of action sets. The set $\mathcal{A}_1$, is pointy, namely, $|\text{Supp}(\mathcal{A}_1)| = 1$, which allows probing only single nodes. The set $\mathcal{A}_2$ is diffuse where only a connected subset of nodes can be queried. In either case, the observation received on an action is as described in Sec. 3.

Key Findings. Here we report pointy actions and isolated anomalies, and other results in Appendix. We choose $\sigma^2 = 0.5$, β such that the false alarms are about 1% for the Oracle, and a forced exploration rate $\epsilon = 0.2$. Our results are averaged over 5000 Monte-Carlo runs. With this choice for β we did not observe false alarms in any of the algorithms.
Gains from Adaptivity. The fact that we perform substantially better than URS uniformly across all of the experiments demonstrates our gains through adaptive sampling.
Data Inefficiency. The fact that the proposed method closely tracks ϵ-GCD+ suggests that our decoupled scheme is capable of fully leveraging the observations, and the cost of decoupled exploration is small.
Optimality. The fact that the proposed method tracks Oracle qualitatively across all experiment setups indicates that our method is close to optimal. This is because through adaptivity we achieve performance similar to a competitor who requires no adaptation (recall that oracle knows the change parameters and thus does not need to explore). In point of fact we note in passing that any other method would at best lie between our proposed method and the Oracle.
Variation with Changepoint ν. For a fixed graph size, we observed that the average expected delay is relatively constant for all methods, which is consistent with Theorem 3.

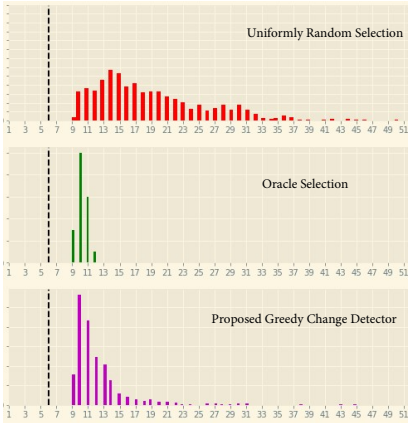

Figure 2: Audio based change detection of machine anomaly: Histogram of stopping times by URS, Oracle, and ϵ-GCD. Histograms for Oracle and ϵ-GCD demonstrates clear adaptivity gains.

Audio based recognition of machine anomalies. We experiment using the MIMII audio dataset [Pur+19b]. Detailed specifics are in the Appendix. The dataset has four machines (ID00,ID02,ID04,ID06), each equipped with audio sensors recording the health of the machine. There are three types of anomaly – rail damage, loose belt, no grease – which can occur at any time, in any machine. Corresponding to each anomaly there is an audio stream, and the anomaly occurs in one of the four machines at an arbitrary time. For each machine, the dataset contains audio-streams of 1000 normal and 300 abnormal files, and each audio-stream is about 10 seconds long.

Audio Processing. For each audio-stream we train auto-encoders on normal data using mel-spectrogram features, and fit Gaussians to the reconstruction errors. This results in pre- and post change parameters for each machine's autoencoded reconstruction error score, corresponding to normal and abnormal operation.

Experiment. To simulate BQCD, we introduced anomalies in machine bearing ID00 as follows. We concatenated 6 normal files and 54 abnormal files chosen uniformly at random from machine ID00. For the other machines we concatenated 60 normal files at random. The 60 files correspond to 600 seconds. Our changepoint corresponds to 6th file, which we denote as $\nu = 6$ and our task is to detect this change. Note that both the machine ID and the changepoint is not known to the learner. Our results are depicted as histograms for changepoints of anomaly detection in Fig. 2. As observed, we notice that ϵ-GCD's performance is close to Oracle both in mean and distribution, while URS exhibits larger delay and significant variance. The appendix simulates histograms for a larger changepoint.

Acknowledgments and Disclosure of Funding

A. G. and B. L. were supported by the Aerospace Network Research Consortium (ANRC) Grant on Airplane IOT Data Management. V. S. was supported by the Army Research Office Grant W911NF2110246, and National Science Foundation grants CCF-2007350 and CCF-1955981. The authors are grateful to Rajesh Sundaresan, Himanshu Tyagi and Manjunath Krishnapur for helpful discussions, and to the anonymous program committee members for their valuable feedback.

References

[AC17] Samaneh Aminikhanghahi and Diane J. Cook. "A Survey of Methods for Time Series Change Point Detection". In: *Knowl. Inf. Syst.* 51.2 (May 2017), pp. 339–367.

[AM07] Ryan Prescott Adams and David J. C. MacKay. *Bayesian Online Changepoint Detection*. 2007. arXiv: `0710.3742 [stat.ML]`.

[AMF20] Reda Alami, Odalric Maillard, and Raphael Feraud. "Restarted Bayesian Online Changepoint Detector achieves Optimal Detection Delay". In: *Proceedings of the 37th International Conference on Machine Learning*. Proceedings of Machine Learning Research. 2020.

[Bas99] Tim Bass. "Multisensor Data Fusion for Next Generation Distributed Intrusion Detection Systems". In: *In Proceedings of the IRIS National Symposium on Sensor and Data Fusion*. 1999, pp. 24–27.

[BN93] Michèle Basseville and Igor V. Nikiforov. *Detection of Abrupt Changes: Theory and Application*. USA: Prentice-Hall, Inc., 1993. ISBN: 0131267809.

[CG12] Jie Chen and Arjun K Gupta. *Parametric Statistical Change Point Analysis: With Applications to Genetics, Medicine, and Finance; 2nd ed.* Boston: Springer, 2012.

[Cov99] Thomas M Cover. *Elements of information theory*. John Wiley & Sons, 1999.

[Din+06] Y. Ding, E.A. Elsayed, S. Kumara, J.-C. Lu, F. Niu, and J. Shi. "Distributed Sensing for Quality and Productivity Improvements". In: *IEEE Transactions on Automation Science and Engineering* (2006).

[ES10] Erhan Baki Ermis and Venkatesh Saligrama. "Distributed Detection in Sensor Networks With Limited Range Multimodal Sensors". In: *IEEE Transactions on Signal Processing* 58.2 (2010), pp. 843–858.

[FL07] Paul Fearnhead and Zhen Liu. "On-line inference for multiple changepoint problems". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 69.4 (Sept. 2007), pp. 589–605.

[GK16] Aurelien Garivier and Emilie Kaufmann. "Optimal Best Arm Identification with Fixed Confidence". In: *Conference On Learning Theory*. 2016, pp. 998–1027.

[GM11] Aurélien Garivier and Eric Moulines. "On upper-confidence bound policies for switching bandit problems". In: *International Conference on Algorithmic Learning Theory*. Springer. 2011, pp. 174–188.

[GMS19] Aurélien Garivier, Pierre Ménard, and Gilles Stoltz. "Explore first, exploit next: The true shape of regret in bandit problems". In: *Mathematics of Operations Research* 44.2 (2019), pp. 377–399.

[Gun+21] Gregory W Gundersen, Diana Cai, Chuteng Zhou, Barbara E Engelhardt, and Ryan P Adams. "Active multi-fidelity Bayesian online changepoint detection". In: *arXiv preprint arXiv:2103.14224* (2021).

[HC11] Alfred O. Hero and Douglas Cochran. "Sensor Management: Past, Present, and Future". In: *IEEE Sensors Journal* (2011).

[HKK19] Shogo Hayashi, Yoshinobu Kawahara, and Hisashi Kashima. "Active Change-Point Detection". In: *Proceedings of The Eleventh Asian Conference on Machine Learning*. Ed. by Wee Sun Lee and Taiji Suzuki. Vol. 101. Proceedings of Machine Learning Research. Nagoya, Japan: PMLR, 17–19 Nov 2019, pp. 1017–1032.

[Jam+14] Kevin Jamieson, Matthew Malloy, Robert Nowak, and Sébastien Bubeck. "lil'ucb: An optimal exploration algorithm for multi-armed bandits". In: *Conference on Learning Theory*. PMLR. 2014, pp. 423–439.

[Lai98] Tze Leung Lai. "Information bounds and quick detection of parameter changes in stochastic systems". In: *IEEE Transactions on Information Theory* 44.7 (1998), pp. 2917–2929.

[LLS18] Fang Liu, Joohyun Lee, and Ness Shroff. "A change-detection based framework for piecewise-stationary multi-armed bandit problem". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 32. 1. 2018.

[Lor71] Gary Lorden. "Procedures for reacting to a change in distribution". In: *The Annals of Mathematical Statistics* 42.6 (1971), pp. 1897–1908.

[LS20] Tor Lattimore and Csaba Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.

[LX10] Tze Leung Lai and Haipeng Xing. "Sequential Change-Point Detection When the Pre- and Post-Change Parameters are Unknown". In: *Sequential Analysis* 29.2 (2010), pp. 162–175.

[Mai19] Odalric-Ambrym Maillard. "Sequential change-point detection: Laplace concentration of scan statistics and non-asymptotic delay bounds". In: *Proceedings of the 30th International Conference on Algorithmic Learning Theory*. Ed. by Aurélien Garivier and Satyen Kale. Vol. 98. Proceedings of Machine Learning Research. Chicago, Illinois: PMLR, 22–24 Mar 2019, pp. 610–632.

[MS13] Joseph Mellor and Jonathan Shapiro. "Thompson sampling in switching environments with Bayesian online change detection". In: *Artificial Intelligence and Statistics*. PMLR. 2013, pp. 442–450.

[OGR10] Michael A. Osborne, Roman Garnett, and Stephen J. Roberts. "Active Data Selection for Sensor Networks with Faults and Changepoints". In: *24th IEEE International Conference on Advanced Information Networking and Applications, AINA 2010, Perth, Australia, 20-13 April 2010*. IEEE Computer Society, 2010, pp. 533–540.

[Pag54] E. S. Page. "Continuous Inspection Schemes". In: *Biometrika* 41.1/2 (1954), pp. 100–115.

[Pur+19a] Harsh Purohit, Ryo Tanabe, Kenji Ichige, Takashi Endo, Yuki Nikaido, Kaori Suefusa, and Yohei Kawaguchi. *MIMII Dataset: Sound Dataset for Malfunctioning Industrial Machine Investigation and Inspection*. 2019. arXiv: `1909.09347`.

[Pur+19b] Harsh Purohit, Ryo Tanabe, Kenji Ichige, Takashi Endo, Yuki Nikaido, Kaori Suefusa, and Yohei Kawaguchi. *MIMII Dataset: Sound Dataset for Malfunctioning Industrial Machine Investigation and Inspection*. Version public 1.0. Zenodo, Sept. 2019. DOI: `10.5281/zenodo.3384388`. URL: `https://doi.org/10.5281/zenodo.3384388`.

[QSC14] Jing Qian, Venkatesh Saligrama, and Yuting Chen. "Connected Sub-graph Detection". In: *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics*. Ed. by Samuel Kaski and Jukka Corander. Vol. 33. Proceedings of Machine Learning Research. Reykjavik, Iceland: PMLR, 22–25 Apr 2014, pp. 796–804.

[Tho33] William R Thompson. "On the likelihood that one unknown probability exceeds another in view of the evidence of two samples". In: *Biometrika* 25.3/4 (1933), pp. 285–294.

[TNB14] Alexander Tartakovsky, Igor Nikiforov, and Michele Basseville. *Sequential analysis: Hypothesis testing and changepoint detection*. CRC Press, 2014.

[VB14] Venugopal V Veeravalli and Taposh Banerjee. "Quickest change detection". In: *Academic Press Library in Signal Processing*. Vol. 3. Elsevier, 2014, pp. 209–255.

[Vil+17] Mauricio Villarroel, João Jorge, Chris Pugh, and Lionel Tarassenko. "Non-Contact Vital Sign Monitoring in the Clinic". In: *2017 12th IEEE International Conference on Automatic Face Gesture Recognition (FG 2017)*. 2017.

[Vis+14] Bimal Viswanath, Muhammad Ahmad Bashir, Mark Crovella, Saikat Guha, Krishna P. Gummadi, Balachander Krishnamurthy, and Alan Mislove. "Towards Detecting Anomalous User Behavior in Online Social Networks". In: *USENIX Security Symposium*. 2014.

[YSK15] Shihao Yang, Mauricio Santillana, and S. C. Kou. "Accurate estimation of influenza epidemics using Google search data via ARGO". In: *Proceedings of the National Academy of Sciences* (2015).

[ZM20] Wanrong Zhang and Yajun Mei. *Bandit Change-Point Detection for Real-Time Monitoring High-Dimensional Data Under Sampling Control*. 2020. arXiv: `2009.11891 [stat.ME]`.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes] See Section 1.
   (b) Did you describe the limitations of your work? [Yes] See Section 5.1 for assumptions under which the results hold.
   (c) Did you discuss any potential negative societal impacts of your work? [No] As a primarily theoretical paper concerned with efficiency, our work does not have any meaningful negative impacts on society beyond the larger issues of how it is employed.
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [Yes] See Section 5.1.
   (b) Did you include complete proofs of all theoretical results? [Yes] See the supplementary material for all proofs.
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes] We have included all the code and instructions to run it as part of the supplementary material. The dataset (MIMII audio sounds) used is large, and is available publicly.
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes] We have specified the details of the experiments and parameter choices in the supplementary material.
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes] We report either standard deviations or empirical distributions in our experiments.
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes] See Section 6.
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [Yes] See subsection on "Audio based recognition of machine anomalies" in Section 6.
   (b) Did you mention the license of the assets? [Yes] See the citation to the MIMII audio dataset in Section 6.
   (c) Did you include any new assets either in the supplemental material or as a URL? [N/A] No new assets were created.
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A] We have not used data from people.
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A] The only measured data used in the paper are those of machine sounds.
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]

Appendix

A Proof of Theorem 2

We in fact establish the following more general result.

Theorem 4. *Let $0 < \alpha \leq \frac{1}{10}$, $m \geq 1$ and $\nu \in \mathbb{N}$. If a bandit changepoint algorithm satisfies $\mathbb{P}^{(\infty)}\left[\tau < \nu + m \;\middle|\; \nu \leq \tau\right] \leq \alpha$, then for any $\theta^* \in \Theta^{(1)}$,*

$$\mathbb{E}^{(\nu,\theta^*)}\left[\tau - \nu \;\middle|\; \tau \geq \nu\right] \geq \min\left\{\frac{\frac{1}{20}\log\frac{1}{\alpha}}{\max_{a\in\mathcal{A}} D\left(\theta^*(a)||\theta_0(a)\right)}, \frac{m}{2}\right\}.$$

Note that Theorem 2 is the special case[8] of $\nu = 1$.

We first recall and/or put down some preliminaries before embarking upon the proof.

Definition: Problem instance. For any changepoint time $\nu \in \{1, 2, \ldots\} \cup \{\infty\}$ and post-change parameter $\theta \in \Theta^{(1)}$, we call the pair (ν, θ) an *instance* of the bandit changepoint detection problem. Note that if $\nu = \infty$, then it is immaterial what the post-change parameter θ is, since there is effectively no change in the distribution; thus we will omit θ, or write $(\infty, *)$, if $\nu = \infty$ for ease of notation. The instance, along with the sampling algorithm and (known) $\theta^{(0)}$, completely determines the distribution of trajectories generated by the operation of the algorithm.

Bandit changepoint detection algorithm: Recall the definition of an algorithm for the bandit changepoint detection problem: It is a rule that maps the history I_t of actions and observations to (1) a decision $U_t \in \{0, 1\}$ to stop playing actions, and (2) if not stopping ($U_t = 0$), then an action A_t to play in round t. Here,

$$I_t = (V_0, U_0, A_1, X_1, V_1, U_1, A_2, X_2, V_2, \ldots, V_{t-2}, U_{t-2}, A_{t-1}, X_{t-1}, V_t),$$

where at any time instant $s \geq 0$, V_s represents independent, internal randomness available to the algorithm, U_s is an indicator random variable for the event that the algorithm decides to stop playing before taking the $(s+1)$-st action (i.e., after playing s actions), and X_s is the observation from playing arm A_s in round s.

Proof of Theorem 4. We first establish an auxiliary lemma about the explicit form for the divergence of the conditional distribution of a trajectory.

Let $\mathrm{Law}_E^{(i,\theta)}(I_\tau)$ denote the probability distribution of the random trajectory I_τ conditioned on the event E, when the algorithm is run on the instance (i, θ).

Lemma 1. *For any parameter $\theta \in \Theta^{(1)}$, $\eta \in \mathbb{N}$ and $E \in \sigma(I_\eta)$,*

$$D\left(Law_E^{(\eta,\theta)}(I_\tau)||Law_E^{(\infty)}(I_\tau)\right) = \sum_{a\in\mathcal{A}} D\left(\theta(a)||\theta_0(a)\right)\mathbb{E}^{(\eta,\theta)}\left[\left(N_\tau(a) - N_{\eta-1}(a)\right)^+ \;\middle|\; E\right].$$

Proof. For the sake of convenience we show the argument assuming that the trajectory I_τ is a discrete random object, i.e., it has a probability mass function. (It is straightforward, but notationally heavier, to extend it to the case of general measures by using Radon-Nikodym derivatives.)

[8] We assume that $\tau \geq 1$ with probability 1.

We have

$$
\begin{aligned}
&D\left(\mathrm{Law}_E^{(\eta,\theta)}(I_\tau)||\mathrm{Law}_E^{(\infty)}(I_\tau)\right)\\
&= \sum_{\omega\equiv(v_0,u_0,a_1,x_1,v_1,\ldots,v_{\tau-1},u_\tau=1)\in E} \mathbb{P}^{(\eta,\theta)}\left[I_\tau=\omega \;\middle|\; E\right]\log\frac{\mathbb{P}^{(\eta,\theta)}\left[I_\tau=\omega \;\middle|\; E\right]}{\mathbb{P}^{(\infty)}\left[I_\tau=\omega|E\right]}\\
&\overset{(a)}{=} \sum_{\omega\equiv(v_0,u_0,a_1,x_1,v_1,\ldots,v_{\tau-1},u_\tau=1)\in E} \mathbb{P}^{(\eta,\theta)}\left[I_\tau=\omega \;\middle|\; E\right]\log\frac{\mathbb{P}^{(\eta,\theta)}\left[I_\tau=\omega\right]}{\mathbb{P}^{(\infty)}\left[I_\tau=\omega\right]}\\
&= \sum_{\omega\in E} \mathbb{P}^{(\eta,\theta)}\left[I_\tau=\omega \;\middle|\; E\right]\log\frac{\mathbb{P}^{(\eta,\theta)}\left[v_0\right]\mathbb{P}^{(\eta,\theta)}\left[u_0|v_0\right]\mathbb{P}^{(\eta,\theta)}\left[a_1|u_0,v_0\right]\mathbb{P}^{(\eta,\theta)}\left[x_1|a_1\right]\cdots}{\mathbb{P}^{(\infty)}\left[v_0\right]\mathbb{P}^{(\infty)}\left[u_0|v_0\right]\mathbb{P}^{(\infty)}\left[a_1|u_0,v_0\right]\mathbb{P}^{(\infty)}\left[x_1|a_1\right]\cdots}\\
&\overset{(b)}{=} \sum_{\omega\in E} \mathbb{P}^{(\eta,\theta)}\left[I_\tau=\omega \;\middle|\; E\right]\log\frac{\mathbb{P}^{(\eta,\theta)}\left[x_1|a_1\right]\mathbb{P}^{(\eta,\theta)}\left[x_2|a_2\right]\cdots}{\mathbb{P}^{(\infty)}\left[x_1|a_1\right]\mathbb{P}^{(\infty)}\left[x_2|a_2\right]\cdots}\\
&= \sum_{\omega\in E} \mathbb{P}^{(\eta,\theta)}\left[I_\tau=\omega \;\middle|\; E\right]\sum_{t=1}^{\tau-1}\log\frac{\mathbb{P}^{(\eta,\theta)}\left[x_t(\omega)|a_t(\omega)\right]}{\mathbb{P}^{(\infty)}\left[x_t(\omega)|a_t(\omega)\right]}\\
&\overset{(c)}{=} \frac{1}{\mathbb{P}^{(\infty)}\left[E\right]}\mathbb{E}^{(\eta,\theta)}\left[\sum_{t=1}^{\tau-1}\log\frac{\mathbb{P}^{(\eta,\theta)}\left[X_t|A_t\right]}{\mathbb{P}^{(\infty)}\left[X_t|A_t\right]}\mathbf{1}\left\{E\right\}\right].
\end{aligned}
$$

Here, (a) and (c) both follow because $\mathbb{P}^{(\eta,\theta)}\left[E\right]=\mathbb{P}^{(\infty)}\left[E\right]$ due to $E\in\sigma(I_\eta)$, and (b) is because the algorithm's decisions and internal randomness v_0, u_0, a_1, a_2, etc. do not depend on the probability distribution of the environment generating the observations. We continue further, writing

$$
\begin{aligned}
&\mathbb{P}^{(\infty)}\left[E\right]\cdot D\left(\mathrm{Law}_E^{(\eta,\theta)}(I_\tau)||\mathrm{Law}_E^{(\infty)}(I_\tau)\right)\\
&= \mathbb{E}^{(\eta,\theta)}\left[\mathbf{1}\left\{E\right\}\sum_{t=1}^{(\eta-1)\wedge(\tau-1)}\log\frac{\mathbb{P}^{(\eta,\theta)}\left[X_t|A_t\right]}{\mathbb{P}^{(\infty)}\left[X_t|A_t\right]}+\mathbf{1}\left\{E\right\}\sum_{t=(\eta-1)\wedge(\tau-1)+1}^{\tau-1}\log\frac{\mathbb{P}^{(\eta,\theta)}\left[X_t|A_t\right]}{\mathbb{P}^{(\infty)}\left[X_t|A_t\right]}\right]\\
&= \mathbb{E}^{(\eta,\theta)}\left[\mathbf{1}\left\{E\right\}\sum_{t=\eta\wedge\tau}^{\tau-1}\log\frac{\mathbb{P}^{(\nu,\theta)}\left[X_t|A_t\right]}{\mathbb{P}^{(\infty)}\left[X_t|A_t\right]}\sum_{a\in\mathcal{A}}\mathbf{1}\left\{A_t=a\right\}\right],
\end{aligned}
$$

because $E\in\sigma(I_\eta)$. Thus,

$$
\begin{aligned}
&\mathbb{P}^{(\infty)}\left[E\right]\cdot D\left(\mathrm{Law}_E^{(\eta,\theta)}(I_\tau)||\mathrm{Law}_E^{(\infty)}(I_\tau)\right)\\
&= \sum_{a\in\mathcal{A}}\sum_{t=\eta\wedge\tau}^{\tau-1}\mathbb{E}^{(\eta,\theta)}\left[\log\frac{\mathbb{P}^{(\nu,\theta)}\left[X_t|A_t\right]}{\mathbb{P}^{(\infty)}\left[X_t|A_t\right]}\mathbf{1}\left\{A_t=a,E\right\}\right]\\
&= \sum_{a\in\mathcal{A}}\sum_{t=\eta\wedge\tau}^{\tau-1}\mathbb{E}^{(\eta,\theta)}\left[\mathbf{1}\left\{A_t=a,E\right\}\mathbb{E}^{(\eta,\theta)}\left[\log\frac{\mathbb{P}^{(\eta,\theta)}\left[X_t|A_t\right]}{\mathbb{P}^{(\infty)}\left[X_t|A_t\right]}\middle|\mathbf{1}\left\{A_t=a,E\right\}\right]\right]\\
&\overset{(d)}{=} \sum_{a\in\mathcal{A}}\sum_{t=\eta\wedge\tau}^{\tau-1}\mathbb{E}^{(\eta,\theta)}\left[\mathbf{1}\left\{A_t=a,E\right\}D\left(\theta(a)||\theta_0(a)\right)\right]\\
&= \sum_{a\in\mathcal{A}}D\left(\theta(a)||\theta_0(a)\right)\mathbb{E}^{(\eta,\theta)}\left[\mathbf{1}\left\{E\right\}\left(N_\tau(a)-N_{\eta-1}(a)\right)^+\right],
\end{aligned}
$$

where (d) is due to $E\in\sigma(I_\eta)$ and the Markov property of the algorithm's trajectory. This completes the proof. □

Returning to the proof of the theorem, we split the analysis into two cases depending on the value of the conditional probability $\mathbb{P}^{(\nu,\theta^*)}\left[\tau<\nu+m \;\middle|\; \nu\le\tau\right]$ of stopping before an additional time m after having crossed the actual changepoint ν.

Case 1: $\mathbb{P}^{(\nu,\theta^*)}\left[\tau < \nu + m \;\middle|\; \nu \le \tau\right] \ge \frac{1}{2}$. In this case, applying the data processing inequality for KL divergence [Cov99] to the two (conditional) input distributions $\text{Law}^{(\nu,\theta^*)}_{\nu\le\tau}(I_\tau) \equiv \mathbb{P}^{(\nu,\theta^*)}\left[I_\tau \in \cdot \;\middle|\; \nu \le \tau\right]$ and $\text{Law}^{(\infty)}_{\nu\le\tau}(I_\tau) \equiv \mathbb{P}^{(\infty)}\left[I_\tau \in \cdot \;\middle|\; \nu \le \tau\right]$, with the data processing function $f(I_\tau) := \mathbf{1}\{\nu \le \tau < \nu + m\}$, yields

$$\begin{aligned}
&D\left(\text{Law}^{(\nu,\theta^*)}_{\nu\le\tau}(I_\tau)||\text{Law}^{(\infty)}_{\nu\le\tau}(I_\tau)\right)\\
&\ge D\left(\text{Ber}\left(\mathbb{P}^{(\nu,\theta^*)}\left[\tau < \nu + m \;\middle|\; \nu \le \tau\right]\right)||\text{Ber}\left(\mathbb{P}^{(\infty)}\left[\tau < \nu + m \;\middle|\; \nu \le \tau\right]\right)\right).
\end{aligned}$$

Together with Lemma 1 for the event $E := \{\nu \le \tau\}$, this gives

$$\begin{aligned}
&\sum_{a\in\mathcal{A}} D\left(\theta^*(a)||\theta_0(a)\right)\mathbb{E}^{(\nu,\theta^*)}\left[(N_\tau(a) - N_{\eta-1}(a))^+ \;\middle|\; \nu \le \tau\right]\\
&\ge D\left(\text{Ber}\left(\mathbb{P}^{(\nu,\theta^*)}\left[\tau < \nu + m \;\middle|\; \nu \le \tau\right]\right)||\text{Ber}\left(\mathbb{P}^{(\infty)}\left[\tau < \nu + m \;\middle|\; \nu \le \tau\right]\right)\right)\\
&\overset{(a)}{\ge} \mathbb{P}^{(\nu,\theta^*)}[\tau < \nu + m \mid \nu \le \tau]\log\frac{\mathbb{P}^{(\nu,\theta^*)}[\tau < \nu + m \mid \nu \le \tau]}{\mathbb{P}^{(\infty)}[\tau < \nu + m \mid \nu \le \tau]} - \ln 2\\
&\overset{(b)}{\ge} \frac{1}{2}\log\frac{1/2}{\alpha} - \ln 2 \ge \frac{1}{20}\log\frac{1}{\alpha},
\end{aligned}$$

by Lemma 2, where (a) is due to [GMS19] and (b) is by hypothesis. We now divide both sides by $\sum_{a\in\mathcal{A}}\mathbb{E}^{(\nu,\theta^*)}\left[(N_\tau(a) - N_{\eta-1}(a))^+ \;\middle|\; \nu \le \tau\right] = \mathbb{E}^{(\nu,\theta^*)}\left[(\tau - \nu)^+ \;\middle|\; \nu \le \tau\right]$, and use the fact that the maximum is at least a convex combination, to get

$$\max_{a\in\mathcal{A}} D\left(\theta^*(a)||\theta_0(a)\right) \ge \frac{1}{20}\frac{\log(1/\alpha)}{\mathbb{E}^{(\nu,\theta^*)}\left[(\tau - \nu)^+ \;\middle|\; \nu \le \tau\right]},$$

giving one part of the theorem.

Case 2: $\mathbb{P}^{(\nu,\theta^*)}\left[\tau < \nu + m \;\middle|\; \nu \le \tau\right] < \frac{1}{2}$. In this case, we have

$$\mathbb{E}^{(\nu,\theta^*)}\left[(\tau - \nu)^+ \;\middle|\; \nu \le \tau\right] \ge m \cdot \mathbb{P}^{(\nu,\theta^*)}\left[\tau \ge \nu + m \;\middle|\; \nu \le \tau\right] \ge \frac{m}{2},$$

giving the other part of the theorem.

Lemma 2. *For* $0 < x < \frac{1}{10}$, *we have* $\frac{1}{2}\log\frac{1}{2x} - \log 2 \ge \frac{1}{20}\log\frac{1}{x}$.

Proof. The proof is by basic calculus. □

□

B Proof of Theorem 3

B.1 Time to false alarm

The ϵ-GCD algorithm stops at the first time t when the largest exponentiated 'exploitation queue' CUSUM statistic, i.e.,

$$\max_{\theta\in\Theta^{(1)}}\exp\left(Q^{(2)}_t(\theta)\right) := \max_{\theta\in\Theta^{(1)}}\max_{1\le s\le t}\prod_{\ell=s}^{t-1}\left(\frac{f_\theta(X_\ell|A_\ell)}{f_{\theta_0}(X_\ell|A_\ell)}\right)^{(1-E_\ell)},$$

exceeds β.

For $m \in \mathbb{N}$, let us compute

$$\begin{aligned}
\mathbb{P}^{(\infty)}[\tau < m] &= \mathbb{P}^{(\infty)}\left[\exists t < m : \max_{\theta \in \Theta^{(1)}} \max_{1 \le s \le t} \prod_{\ell=s}^{t-1} \left(\frac{f_\theta(X_\ell|A_\ell)}{f_{\theta_0}(X_\ell|A_\ell)}\right)^{(1-E_\ell)} \ge \beta\right] \\
&\le \mathbb{P}^{(\infty)}\left[\exists \theta \in \Theta^{(1)}, s \in [m] : \max_{s \le t \le m} \prod_{\ell=s}^{t-1} \left(\frac{f_\theta(X_\ell|A_\ell)}{f_{\theta_0}(X_\ell|A_\ell)}\right)^{(1-E_\ell)} \ge \beta\right] \\
&\le \sum_{\theta \in \Theta^{(1)}} \sum_{s=1}^{m} \mathbb{P}^{(\infty)}\left[\max_{s \le t \le m} \prod_{\ell=s}^{t-1} \left(\frac{f_\theta(X_\ell|A_\ell)}{f_{\theta_0}(X_\ell|A_\ell)}\right)^{(1-E_\ell)} \ge \beta\right], \qquad (2)
\end{aligned}$$

by a union bound.

Lemma 3. *For each fixed $s \in [m]$ and $\theta \in \Theta^{(1)}$, the likelihood ratio process $\left\{\prod_{\ell=s}^{t-1} \left(\frac{f_\theta(X_\ell|A_\ell)}{f_{\theta_0}(X_\ell|A_\ell)}\right)^{(1-E_\ell)}\right\}_{t \ge s}$ is a mean-1 martingale under the measure $\mathbb{P}^{(\infty)}$ and with respect to the filtration $(\mathcal{F}_t)_{t \ge s}$, where*

$$\forall t \ge s \quad \mathcal{F}_t := \sigma(E_s, A_s, X_s, E_{s+1}, A_{s+1}, X_{s+1}, \ldots, E_{t-1}, A_{t-1}, X_{t-1}).$$

Proof. Taking the conditional expectation of the t-th term of the process w.r.t. $\mathcal{F}_{t-1}$, we get

$$\begin{aligned}
&\mathbb{E}^{(\infty)}\left[\prod_{\ell=s}^{t-1} \left(\frac{f_\theta(X_\ell|A_\ell)}{f_{\theta_0}(X_\ell|A_\ell)}\right)^{(1-E_\ell)} \middle| \mathcal{F}_{t-1}\right] \\
&= \prod_{\ell=s}^{t-2} \left(\frac{f_\theta(X_\ell|A_\ell)}{f_{\theta_0}(X_\ell|A_\ell)}\right)^{(1-E_\ell)} \mathbb{E}^{(\infty)}\left[\left(\frac{f_\theta(X_{t-1}|A_{t-1})}{f_{\theta_0}(X_{t-1}|A_{t-1})}\right)^{(1-E_{t-1})} \middle| \mathcal{F}_{t-1}\right].
\end{aligned}$$

The conditional expectation on the right hand side satisfies

$$\begin{aligned}
&\mathbb{E}^{(\infty)}\left[\left(\frac{f_\theta(X_{t-1}|A_{t-1})}{f_{\theta_0}(X_{t-1}|A_{t-1})}\right)^{(1-E_{t-1})} \middle| \mathcal{F}_{t-1}\right] \\
&= \mathbb{P}^{(\infty)}[E_{t-1} = 1] + \mathbb{P}^{(\infty)}[E_{t-1} = 0] \cdot \mathbb{E}^{(\infty)}\left[\frac{f_\theta(X_{t-1}|A_{t-1})}{f_{\theta_0}(X_{t-1}|A_{t-1})} \middle| \mathcal{F}_{t-1}\right],
\end{aligned}$$

where we have used the independence of the exploration decision E_{t-1} from the past. By iterated expectation, we have

$$\begin{aligned}
\mathbb{E}^{(\infty)}\left[\frac{f_\theta(X_{t-1}|A_{t-1})}{f_{\theta_0}(X_{t-1}|A_{t-1})} \middle| \mathcal{F}_{t-1}\right] &= \mathbb{E}^{(\infty)}\left[\mathbb{E}^{(\infty)}\left[\frac{f_\theta(X_{t-1}|A_{t-1})}{f_{\theta_0}(X_{t-1}|A_{t-1})} \middle| A_{t-1}, \mathcal{F}_{t-1}\right] \middle| \mathcal{F}_{t-1}\right] \\
&= \mathbb{E}^{(\infty)}\left[1 \mid \mathcal{F}_{t-1}\right],
\end{aligned}$$

establishing the result. □

Thanks to Lemma 3, we can apply Doob's maximal inequality to the non-negative likelihood ratio martingale above to get

$$\mathbb{P}^{(\infty)}\left[\max_{s \le t \le m} \prod_{\ell=s}^{t-1} \left(\frac{f_\theta(X_\ell|A_\ell)}{f_{\theta_0}(X_\ell|A_\ell)}\right)^{(1-E_\ell)} \ge \beta\right] \le \frac{1}{\beta}.$$

Together with (2), this implies

$$\mathbb{P}^{(\infty)}[\tau < m] \le \sum_{\theta \in \Theta^{(1)}} \sum_{s=1}^{m} \mathbb{P}^{(\infty)}\left[\max_{s \le t \le m} \prod_{\ell=s}^{t-1} \left(\frac{f_\theta(X_\ell|A_\ell)}{f_{\theta_0}(X_\ell|A_\ell)}\right)^{(1-E_\ell)} \ge \beta\right] \le \frac{\left|\Theta^{(1)}\right| m}{\beta}.$$

Therefore, using a stopping threshold satisfying $\beta \ge \frac{m\left|\Theta^{(1)}\right|}{\alpha}$ guarantees the false alarm property $\mathbb{P}^{(\infty)}[\tau < m] \le \alpha$.

B.2 Detection delay

Preliminaries. Let the true post-change parameter starting from an arbitrary time $\nu \in \mathbb{N}$ be equal to $\theta^* \in \Theta^{(1)}$.

Recall that the ϵ-GCD algorithm (Algorithm 1) makes, at round $t \geq 1$, the generalized maximum likelihood estimate

$$\begin{aligned}
\hat{\theta}_t &= \operatorname*{argmax}_{\theta \in \Theta^{(1)}} \max_{v=1}^{t} \prod_{\ell=1}^{v-1} f_{\theta_0}(X_\ell|A_\ell)^{E_\ell} \prod_{\ell=v}^{t-1} f_\theta(X_\ell|A_\ell)^{E_\ell} \\
&= \operatorname*{argmax}_{\theta \in \Theta^{(1)}} \max_{v=1}^{t} \prod_{\ell=1}^{v-1} f_{\theta_0}(X_\ell|A_\ell)^{E_\ell} \prod_{\ell=v}^{t-1} f_\theta(X_\ell|A_\ell)^{E_\ell} \Big/ \prod_{\ell=1}^{t-1} f_{\theta_0}(X_\ell|A_\ell)^{E_\ell} \\
&= \operatorname*{argmax}_{\theta \in \Theta^{(1)}} \max_{v=1}^{t} \prod_{\ell=v}^{t-1} \left(\frac{f_\theta(X_\ell|A_\ell)}{f_{\theta_0}(X_\ell|A_\ell)} \right)^{E_\ell} = \operatorname*{argmax}_{\theta \in \Theta^{(1)}} \max_{v=1}^{t} J_{v,t-1}(\theta) = \operatorname*{argmax}_{\theta \in \Theta^{(1)}} \max_{v=1}^{t} \log J_{v,t-1}(\theta),
\end{aligned}$$

where we have denoted $J_{v,t-1}(\theta) := \prod_{\ell=v}^{t-1} \left(\frac{f_\theta(X_\ell|A_\ell)}{f_{\theta_0}(X_\ell|A_\ell)} \right)^{E_\ell} \Leftrightarrow \log J_{v,t-1}(\theta) = \sum_{\ell=v}^{t-1} E_\ell \log \frac{f_\theta(X_\ell|A_\ell)}{f_{\theta_0}(X_\ell|A_\ell)} = \sum_{\ell=v}^{t-1} U_\ell(\theta)$, where we have defined $U_\ell(\theta) \equiv U_\ell(\theta, \theta_0, X_\ell, A_\ell) := E_\ell \log \frac{f_\theta(X_\ell|A_\ell)}{f_{\theta_0}(X_\ell|A_\ell)}$. Also recall that $Q_t^{(1)}(\theta) = \max_{v=1}^{t} \log J_{v,t-1}(\theta) \geq 0$ represents the CUSUM-style 'exploration queue' statistic for each candidate parameter $\theta \in \Theta^{(1)}$.

Step 1: Bounding the time for the 'right CUSUM queue $Q_t^{(1)}(\theta^*)$ to outstrip other queues.

Suppose $\theta \in \Theta^{(1)}$ satisfies $a_\theta^* \neq a_{\theta^*}^*$. For an arbitrary time $t \geq \nu$, we can write

$$\begin{aligned}
&\mathbb{P}^{(\nu,\theta^*)} \left[\max_{i=1}^{t} \log J_{i,t-1}(\theta) \geq \max_{i=1}^{t} \log J_{i,t-1}(\theta^*) \right] = \mathbb{P}^{(\nu,\theta^*)} \left[\max_{i=1}^{t} \sum_{\ell=i}^{t-1} U_\ell(\theta) \geq \max_{i=1}^{t} \sum_{\ell=i}^{t-1} U_\ell(\theta^*) \right] \\
&\leq \mathbb{P}^{(\nu,\theta^*)} \left[\max_{i=1}^{\nu} \sum_{\ell=i}^{\nu-1} U_\ell(\theta) + \max_{i=\nu}^{t} \sum_{\ell=i}^{t-1} U_\ell(\theta) \geq \max_{i=1}^{t} \sum_{\ell=i}^{t-1} U_\ell(\theta^*) \right] \quad \text{(by Lemma 4)} \\
&\leq \mathbb{P}^{(\nu,\theta^*)} \left[\max_{i=1}^{\nu} \sum_{\ell=i}^{\nu-1} U_\ell(\theta) + \max_{i=\nu}^{t} \sum_{\ell=i}^{t-1} U_\ell(\theta) \geq \sum_{\ell=\nu}^{t-1} U_\ell(\theta^*) \right] \quad \text{(since } \nu \in [t]\text{)} \\
&= \mathbb{P}^{(\nu,\theta^*)} \left[Q_\nu^{(1)}(\theta) + \max_{i=\nu}^{t} \sum_{\ell=i}^{t-1} U_\ell(\theta) \geq \sum_{\ell=\nu}^{t-1} U_\ell(\theta^*) \right] \\
&= \mathbb{E}^{(\nu,\theta^*)} \left[\mathbb{P}^{(\nu,\theta^*)} \left[Q_\nu^{(1)}(\theta) + \max_{i=\nu}^{t} \sum_{\ell=i}^{t-1} U_\ell(\theta) \geq \sum_{\ell=\nu}^{t-1} U_\ell(\theta^*) \;\middle|\; Q_\nu^{(1)}(\theta) \right] \right]. \\
&= \mathbb{E}^{(\infty)} \left[\mathbb{P}^{(\nu,\theta^*)} \left[Q_\nu^{(1)}(\theta) + \max_{i=\nu}^{t} \sum_{\ell=i}^{t-1} U_\ell(\theta) \geq \sum_{\ell=\nu}^{t-1} U_\ell(\theta^*) \;\middle|\; Q_\nu^{(1)}(\theta) \right] \right]. \qquad (3)
\end{aligned}$$

The final equality is due to the fact that the inner conditional probability is a function of only $Q_\nu^{(1)}(\theta)$, whose distribution is identical under $\mathbb{P}^{(\nu,\theta^*)}$ and $\mathbb{P}^{(\infty)}$ because it is determined by actions and observations before the change time ν.

We now make the crucial observation that for any $q \geq 0$,

$$\begin{aligned}
&\mathbb{P}^{(\nu,\theta^*)} \left[Q_\nu(\theta) + \max_{i=\nu}^{t} \sum_{\ell=i}^{t-1} U_\ell(\theta) \geq \sum_{\ell=\nu}^{t-1} U_\ell(\theta^*) \;\middle|\; Q_\nu(\theta) = q \right] \\
&= \mathbb{P}^{(\nu,\theta^*)} \left[q + \max_{i=\nu}^{t} \sum_{\ell=i}^{t-1} U_\ell(\theta) \geq \sum_{\ell=\nu}^{t-1} U_\ell(\theta^*) \;\middle|\; Q_\nu(\theta) = q \right] \\
&= \mathbb{P}^{(1,\theta^*)} \left[q + \max_{i=1}^{t-\nu+1} \sum_{\ell=i}^{t-\nu} U_\ell(\theta) \geq \sum_{\ell=1}^{t-\nu} U_\ell(\theta^*) \right]. \qquad (4)
\end{aligned}$$

The first equality above is by simply substituting for $Q_\nu(\theta)$, but the second equality holds because (a) the random variables E_ℓ, X_ℓ, A_ℓ for $\ell \geq \nu$ are independent of $Q_\nu(\theta)$, by virtue of the independent forced exploration enforced in the algorithm, and (b) the probability distribution of exploration actions and their corresponding observations from round ν onward under $\mathbb{P}^{(\nu,\theta^*)}$ is the same as that of the observations and actions from round 1 onward under $\mathbb{P}^{(1,\theta^*)}$. In other words, *we have rewound the time clock so that the change point is at time* 1 *instead of time* $\nu \geq 1$.

Going forward, to lighten notation, we use $\mathbb{E}$ and $\mathbb{P}$ instead of $\mathbb{E}^{(1,\theta^*)}$ and $\mathbb{P}^{(1,\theta^*)}$ in our calculations. We start by bounding the expectation on the right hand side of (4):

$$\begin{aligned}
&\mathbb{P}\left[q + \max_{i=1}^{t-\nu+1} \sum_{\ell=i}^{t-\nu} U_\ell(\theta) \geq \sum_{\ell=1}^{t-\nu} U_\ell(\theta^*)\right] \\
&\leq \mathbb{P}\left[\psi \geq \sum_{\ell=1}^{t-\nu} U_\ell(\theta^*)\right] + \mathbb{P}\left[q + \max_{i=1}^{t-\nu+1} \sum_{\ell=i}^{t-\nu} U_\ell(\theta) \geq \psi\right], \qquad (5)
\end{aligned}$$

where ψ is chosen as follows. We first introduce the shorthand $D_{\theta_1,\theta_2}(a) := D(\theta_1(a)||\theta_2(a))$. We then define, for each $\theta \in \Theta^{(1)} \cup \{\theta_0\}$,

$$\bar{D}_{\theta^*,\theta} := \mathbb{E}^{(1,\theta^*)}\left[D_{\theta^*,\theta}(A_\ell) \;\middle|\; E_\ell = 1\right] = \sum_{a\in\mathcal{A}} \pi(a) D_{\theta^*,\theta}(a)$$

to be the average KL divergence between θ^* and θ, when an action is randomly sampled from the exploration distribution π. Intuitively, $\epsilon(\bar{D}_{\theta^*,\theta_0} - \bar{D}_{\theta^*,\theta})$ is the average rate of drift of the queue $Q_t^{(1)}(\theta)$ at any time after the changepoint $t \geq \nu$ within the exploration rounds; thus, the queue $Q_t^{(1)}(\theta^*)$ enjoys the highest possible drift rate upward.

Finally, we let

$$\psi := \epsilon(t-\nu) \cdot \frac{\bar{D}_{\theta^*,\theta_0} + \left(\bar{D}_{\theta^*,\theta_0} - \bar{D}_{\theta^*,\theta}\right)^+}{2}.$$

We will also find it useful to define the 'gap' of a parameter θ w.r.t. the true post-change parameter θ^* as

$$\Delta_\theta := \bar{D}_{\theta^*,\theta_0} - \frac{\bar{D}_{\theta^*,\theta_0} + \left(\bar{D}_{\theta^*,\theta_0} - \bar{D}_{\theta^*,\theta}\right)^+}{2} = \frac{1}{2}\min\left(\bar{D}_{\theta^*,\theta_0}, \bar{D}_{\theta^*,\theta}\right).$$

Note that $0 < \Delta_\theta \leq \frac{\bar{D}_{\theta^*,\theta_0}}{2}$ by Assumption 2.

To bound the first term on the right of (5), we introduce the notation $W_\ell(\theta) \equiv W_\ell(\theta, \theta_0, \theta^*, X_\ell, A_\ell) := \log \frac{f_\theta(X_\ell|A_\ell)}{f_{\theta_0}(X_\ell|A_\ell)} + D_{\theta^*,\theta}(A_\ell) - D_{\theta^*,\theta_0}(A_\ell)$. With this we can write

$$\begin{aligned}
&\mathbb{P}\left[\psi \geq \sum_{\ell=1}^{t-\nu} U_\ell(\theta^*)\right] = \mathbb{P}\left[\psi \geq \sum_{\ell=1}^{t-\nu} E_\ell\left(W_\ell(\theta^*) + D_{\theta^*,\theta_0}(A_\ell)\right)\right] \\
&= \mathbb{P}\left[\sum_{\ell=1}^{t-\nu} E_\ell W_\ell(\theta^*) + E_\ell D_{\theta^*,\theta_0}(A_\ell) < \epsilon(t-\nu)(\bar{D}_{\theta^*,\theta_0} - \Delta_\theta)\right] \\
&\leq \exp\left(-\frac{\epsilon^2(t-\nu)^2\Delta_\theta^2}{2(r + D_{\max}^2)(t-\nu)}\right) = \exp\left(-\frac{\epsilon^2(t-\nu)\Delta_\theta^2}{2(r + D_{\max}^2)}\right), \qquad (6)
\end{aligned}$$

by a standard Chernoff bound for sums of iid subgaussian random variables; this is because each iid random variable $E_\ell W_\ell(\theta^*) + E_\ell D_{\theta^*,\theta_0}(A_\ell)$ is subgaussian with (variance) parameter $r + D_{\max}^2$ and mean $\epsilon\bar{D}_{\theta^*,\theta_0}$.

On the other hand, the second term on the right side of (5) can be bounded as follows:

$$\mathbb{P}\left[q+\max_{i=1}^{t-\nu+1}\sum_{\ell=i}^{t-\nu}U_\ell(\theta)\geq\psi\right]=\mathbb{P}\left[\max_{i=1}^{t-\nu+1}\sum_{\ell=i}^{t-\nu}E_\ell\left(W_\ell(\theta)+D_{\theta^*,\theta_0}\left(A_\ell\right)-D_{\theta^*,\theta}\left(A_\ell\right)\right)>\psi-q\right]$$

$$=\mathbb{P}\left[\max_{i=1}^{t-\nu+1}\sum_{\ell=i}^{t-\nu}E_\ell W_\ell(\theta)+\left(E_\ell D_{\theta^*,\theta_0}\left(A_\ell\right)-\epsilon\bar{D}_{\theta^*,\theta_0}\right)+\left(-E_\ell D_{\theta^*,\theta}\left(A_\ell\right)+\epsilon\bar{D}_{\theta^*,\theta}\right)\right.$$

$$\left.+\ \epsilon\bar{D}_{\theta^*,\theta_0}-\epsilon\bar{D}_{\theta^*,\theta}>\psi-q\right].$$

The first three terms of each summand above are zero mean and subgaussian with a total variance parameter of $r+2D_{\max}^2$. Denoting their sum by $D_\ell:=E_\ell W_\ell(\theta)+\left(E_\ell D_{\theta^*,\theta_0}\left(A_\ell\right)-\epsilon\bar{D}_{\theta^*,\theta_0}\right)+\left(-E_\ell D_{\theta^*,\theta}\left(A_\ell\right)+\epsilon\bar{D}_{\theta^*,\theta}\right)$, we split the analysis into two cases.

Case 1. If $\bar{D}_{\theta^*,\theta_0}-\bar{D}_{\theta^*,\theta}>0$, then

$$\begin{aligned}
&\mathbb{P}\left[\max_{i=1}^{t-\nu+1}\sum_{\ell=i}^{t-\nu}(D_\ell+\epsilon\bar{D}_{\theta^*,\theta_0}-\epsilon\bar{D}_{\theta^*,\theta})>\psi-q\right]\\
&\leq\mathbb{P}\left[\max_{i=1}^{t-\nu+1}\sum_{\ell=i}^{t-\nu}D_\ell+\max_{i=1}^{t-\nu+1}\sum_{\ell=i}^{t-\nu}\epsilon(\bar{D}_{\theta^*,\theta_0}-\bar{D}_{\theta^*,\theta})>\psi-q\right]\\
&=\mathbb{P}\left[\max_{i=1}^{t-\nu+1}\sum_{\ell=i}^{t-\nu}D_\ell+\epsilon(t-\nu)(\bar{D}_{\theta^*,\theta_0}-\bar{D}_{\theta^*,\theta})>\epsilon(t-\nu)\left(\bar{D}_{\theta^*,\theta_0}-\frac{\bar{D}_{\theta^*,\theta}}{2}\right)-q\right]\\
&=\mathbb{P}\left[\max_{i=1}^{t-\nu+1}\sum_{\ell=i}^{t-\nu}D_\ell>\epsilon(t-\nu)\Delta_\theta-q\right]\leq\exp\left(-\frac{(\epsilon(t-\nu)\Delta_\theta-q)^2}{2(r+2D_{\max}^2)(t-\nu)}\right),
\end{aligned}\tag{7}$$

thanks to Hoeffding's maximal inequality whenever $\epsilon(t-\nu)\Delta_\theta>q$, see e.g., [Jam+14].

Case 2. If $\bar{D}_{\theta^*,\theta_0}-\bar{D}_{\theta^*,\theta}\leq 0$, then

$$\begin{aligned}
&\mathbb{P}\left[\max_{i=1}^{t-\nu+1}\sum_{\ell=i}^{t-\nu}(D_\ell+\epsilon\bar{D}_{\theta^*,\theta_0}-\epsilon\bar{D}_{\theta^*,\theta})>\psi-q\right]\\
&\leq\mathbb{P}\left[\max_{i=1}^{t-\nu+1}\sum_{\ell=i}^{t-\nu}D_\ell+\max_{v=1}^{t-1}\sum_{\ell=v}^{t-1}\epsilon(\bar{D}_{\theta^*,\theta_0}-\bar{D}_{\theta^*,\theta})>\psi-q\right]\\
&=\mathbb{P}\left[\max_{i=1}^{t-\nu+1}\sum_{\ell=i}^{t-\nu}D_\ell>\epsilon(t-\nu)\Delta_\theta-q\right]\leq\exp\left(-\frac{(\epsilon(t-\nu)\Delta_\theta-q)^2}{2(r+2D_{\max}^2)(t-\nu)}\right),
\end{aligned}\tag{8}$$

again thanks to Hoeffding's maximal inequality whenever $\epsilon(t-\nu)\Delta_\theta>q$.

Define the following 'bad' event B_t, representing the situation that the queue $Q_t^{(1)}(\theta^*)$ has not overtaken some other queue $Q_t^{(1)}(\theta)$ by time t:

$$B_t:=\bigcup_{\theta:a_\theta^*\neq a_{\theta^*}^*}\left\{\max_{v=1}^{t-1}\log J_{v,t-1}(\theta)\geq\max_{v=1}^{t-1}\log J_{v,t-1}(\theta^*)\right\}.$$

Collecting (3)-(8), and employing an additional union bound over $\Theta^{(1)}$, gives $\forall t\geq\nu$:

$$\mathbb{P}^{(\nu,\theta^*)}\left[B_t\mid Q_\nu^{(1)}\right]\leq$$

$$\sum_{\theta:a_\theta^*\neq a_{\theta^*}^*}\left\{\exp\left(-\frac{\epsilon^2(t-\nu)\Delta_\theta^2}{2(r+D_{\max}^2)}\right)+\mathbf{1}\left\{\kappa_{t-\nu}(\theta)^c\right\}+\mathbf{1}\left\{\kappa_{t-\nu}(\theta)\right\}\exp\left(-\frac{\left(\epsilon(t-\nu)\Delta_\theta-Q_\nu^{(1)}(\theta)\right)^2}{2(t-\nu)(r+2D_{\max}^2)}\right)\right\}\tag{9}$$

where we have defined $Q_\nu^{(1)} \equiv \left(Q_\nu^{(1)}(\theta)\right)_{\theta\in\Theta^{(1)}}$ to be the set of all CUSUM exploration statistics across parameters, at the beginning of round ν, and denoted $\kappa_{t-\nu}(\theta) := \left\{Q_\nu^{(1)}(\theta) < \epsilon(t-\nu)\Delta_\theta\right\}$. Note that by the definition of the algorithm, we have $B_t^c \cap \{E_t = 0\} \subseteq \{A_t = a^*_{\theta^*}\}$.

Step 2: Bounding the additional time for the CUSUM queue $Q_t^{(2)}(\theta^*)$ to rise above the threshold and trigger stopping.

For each $\theta \in \Theta^{(1)}$ and $s, t \in \mathbb{N}$, recall the exploitation-based CUSUM statistic for θ_0 versus θ [Lor71], after having accumulated t rounds worth of observations in exploitation phases:

$$Q_{t+1}^{(2)}(\theta) := \log \max_{1\le s\le t+1} \prod_{\ell=s}^{t} \left(\frac{f_\theta(X_\ell|A_\ell)}{f_{\theta_0}(X_\ell|A_\ell)}\right)^{1-E_\ell} = \max_{1\le s\le t+1} \sum_{\ell=s}^{t} (1-E_\ell)\log\left(\frac{f_\theta(X_\ell|A_\ell)}{f_{\theta_0}(X_\ell|A_\ell)}\right),$$

where the empty product is defined to be 1, as usual. This statistic satisfies the recursive relation

$$Q_{t+1}^{(2)}(\theta) = \left(Q_t^{(2)}(\theta) + (1-E_t)\log\left(\frac{f_\theta(X_t|A_t)}{f_{\theta_0}(X_t|A_t)}\right)\right)^+.$$

Moreover, the algorithm stops as soon as $\max_{\theta\in\Theta^{(1)}} Q_t^{(2)}(\theta)$ exceeds the level $\log\beta$.

To lighten our notational burden, we henceforth use $\mathbb{P}_Q^{(\nu,\theta^*)}$ to denote the conditional measure $\mathbb{P}^{(\nu,\theta^*)}[\cdot \mid Q_\nu^{(1)}]$. We have, for any positive integer k, that

$$\mathbb{P}_Q^{(\nu,\theta^*)}\left[(\nu,\theta^*)\right]\tau \ge \nu+k \le \mathbb{P}_Q^{(\nu,\theta^*)}\left[\tau \ge \nu+k, \bigcap_{t=\nu+\frac{k}{2}}^{\nu+k} B_t^c\right] + \mathbb{P}_Q^{(\nu,\theta^*)}\left[\bigcup_{t=\nu+\frac{k}{2}}^{\nu+k} B_t\right]. \tag{10}$$

Also, by the definition of τ and the maximum-of-partial-sums property of $Q_t^{(2)}(\theta)$, we have

$$\begin{aligned}
&\mathbb{P}_Q^{(\nu,\theta^*)}\left[\tau \ge \nu+k, \bigcap_{t=\nu+\frac{k}{2}}^{\nu+k} B_t^c\right] \le \mathbb{P}_Q^{(\nu,\theta^*)}\left[\sum_{\ell=\nu+\frac{k}{2}}^{\nu+k}(1-E_\ell)\log\left(\frac{f_{\theta^*}(X_\ell|A_\ell)}{f_{\theta_0}(X_\ell|A_\ell)}\right) < \log\beta, \bigcap_{t=\nu+\frac{k}{2}}^{\nu+k} B_t^c\right] \\
&\le \mathbb{P}_Q^{(\nu,\theta^*)}\left[\sum_{\ell=\nu+\frac{k}{2}}^{\nu+k}(1-E_\ell)\log\left(\frac{f_{\theta^*}(X_\ell|A_\ell)}{f_{\theta_0}(X_\ell|A_\ell)}\right) < \log\beta, \bigcap_{t=\nu+\frac{k}{2}}^{\nu+k} B_t^c, G_{\nu+k/2,\nu+k}\right] + \mathbb{P}_Q^{(\nu,\theta^*)}\left[G^c_{\nu+k/2,\nu+k}\right],
\end{aligned} \tag{11}$$

where we have defined the 'good' events

$$G_{i,j} := \left\{\sum_{s=i}^{j}(1-E_s) \ge \frac{1}{2}(j-i+1)(1-\epsilon)\right\}$$

for any i, j, representing an adequate amount of exploitation in the time interval $\{i, i+1, \ldots, j\}$. Assuming $\epsilon \le 1/2$, by Hoeffding's inequality we get

$$\mathbb{P}_Q^{(\nu,\theta^*)}\left[G^c_{\nu+k/2,\nu+k}\right] \le e^{-k/16}. \tag{12}$$

By the law of total probability, we have

$$\begin{aligned}
&\mathbb{P}_Q^{(\nu,\theta^*)}\left[\sum_{\ell=\nu+\frac{k}{2}}^{\nu+k}(1-E_\ell)\log\left(\frac{f_{\theta^*}(X_\ell|A_\ell)}{f_{\theta_0}(X_\ell|A_\ell)}\right) < \log\beta, \bigcap_{t=\nu+\frac{k}{2}}^{\nu+k} B_t^c, G_{\nu+k/2,\nu+k}\right] \\
&= \sum_{j=\frac{1}{4}k(1-\epsilon)}^{k/2} \mathbb{P}_Q^{(\nu,\theta^*)}\left[\sum_{s=\nu+k/2}^{\nu+k}(1-E_s) = j, \bigcap_{t=\nu+k/2}^{\nu+k} B_t^c\right] \times \\
&\qquad \mathbb{P}_Q^{(\nu,\theta^*)}\left[\sum_{\ell=\nu+k/2}^{\nu+k}(1-E_\ell)\log\left(\frac{f_{\theta^*}(X_\ell|A_\ell)}{f_{\theta_0}(X_\ell|A_\ell)}\right) < \log\beta \,\middle|\, \sum_{s=\nu+k/2}^{\nu+k}(1-E_s) = j, \bigcap_{t=\nu+k/2}^{\nu+k} B_t^c\right].
\end{aligned} \tag{13}$$

Recall that under $\mathbb{P}_Q^{(\nu,\theta^*)}[\cdot]$, for any $\ell \geq \nu$, the random variable $\log \frac{f_{\theta^*}(X_\ell|A_\ell)}{f_{\theta_0}(X_\ell|A_\ell)}$ is r-subgaussian and has mean $D\left(\theta^*(A_\ell)||\theta_0(A_\ell)\right)$, conditioned on the past trajectory up to and including A_ℓ.

Consider now an alternative (and equivalent) probability space where the sequence of observations from playing the action $a^*_{\theta^*}$ in any exploitation round not earlier than $\nu + k/2$ (i.e., any round index $\ell \geq \nu + k/2$ with $E_\ell = 0$) is revealed sequentially in order from the iid sequence $Z_1, Z_2, \ldots$, where each Z_i has the probability distribution $\mathbb{P}^{(1,\theta^*)}\left[\log \frac{f_{\theta^*}(X|a^*_{\theta^*})}{f_{\theta_0}(X|a^*_{\theta^*})}\right]$. Define $\mu^* := D\left(\theta^*(a^*_{\theta^*})||\theta_0(a^*_{\theta^*})\right)$ to be the mean of each Z_i.

We invoke standard subgaussian concentration of iid sums in this equivalent probability space, say $\tilde{\mathbb{P}}$, to get

$$\begin{aligned}
&\mathbb{P}_Q^{(\nu,\theta^*)}\left[\sum_{\ell=\nu+k/2}^{\nu+k}(1-E_\ell)\log\left(\frac{f_{\theta^*}(X_\ell|A_\ell)}{f_{\theta_0}(X_\ell|A_\ell)}\right) < \log\beta \,\middle|\, \sum_{s=\nu+k/2}^{\nu+k}(1-E_s) = j, \bigcap_{t=\nu+k/2}^{\nu+k} B_t^c\right] \\
&= \tilde{\mathbb{P}}\left[\sum_{i=1}^{j} Z_i < \log\beta \,\middle|\, \sum_{s=\nu+k/2}^{\nu+k}(1-E_s) = j, \bigcap_{t=\nu+k/2}^{\nu+k} B_t^c\right] \\
&= \tilde{\mathbb{P}}\left[\sum_{i=1}^{j} Z_i < \log\beta\right] \quad (\because \{Z_i\}_i \text{ depend only on exploitation outcomes, } \{B_t\}_t \text{ depend only on exploration outcomes}) \\
&\leq \exp\left(-\frac{(\log\beta - j\mu^*)^2}{2jr}\right)
\end{aligned}$$

whenever $j \geq \frac{\log\beta}{\mu^*}$, by a Chernoff bound. Using this in (13) gives

$$\begin{aligned}
&\mathbb{P}_Q^{(\nu,\theta^*)}\left[\sum_{\ell=\nu+k/2}^{\nu+k}(1-E_\ell)\log\left(\frac{f_{\theta^*}(X_\ell|A_\ell)}{f_{\theta_0}(X_\ell|A_\ell)}\right) < \log\beta, \bigcap_{t=\nu+k/2}^{\nu+k} B_t^c, G_{\nu+k/2,\nu+k}\right] \\
&\leq \max_{j=\frac{1}{4}k(1-\epsilon)}^{k/2} \exp\left(-\frac{(\log\beta - j\mu^*)^2}{2jr}\right) = \exp\left(-\frac{(\log\beta - j^*\mu^*)^2}{2j^*r}\right) \qquad (14)
\end{aligned}$$

with $j^* := \frac{k}{4}(1-\epsilon)$, as long as $j^* \geq \frac{\log\beta}{\mu^*} \Leftrightarrow k \geq \frac{4\log\beta}{\mu^*(1-\epsilon)}$. This follows by the fact that the function $x \mapsto \frac{x^2}{a+x}$ with $a > 0$ is increasing in $(0, \infty)$.

Step 3. Putting together the time bounds to get an overall delay bound.

Putting together (9)-(14) and denoting $\gamma := r + D_{\max}^2$, we get that whenever $k \geq \frac{4\log\beta}{\mu^*(1-\epsilon)}$,

$$\begin{aligned}
&\mathbb{P}_Q^{(\nu,\theta^*)}[\tau \geq \nu + k] \\
&\leq \sum_{\theta: a^*_\theta \neq a^*_{\theta^*}} \sum_{t=\nu+k/2}^{\nu+k} \left\{\exp\left(-\frac{\epsilon^2(t-\nu)\Delta_\theta^2}{2\gamma}\right) + \mathbf{1}\left\{\kappa_{t-\nu}(\theta)^c\right\} + \mathbf{1}\left\{\kappa_{t-\nu}(\theta)\right\} e^{-\frac{\left(\epsilon(t-\nu)\Delta_\theta - Q_\nu^{(1)}(\theta)\right)^2}{2(t-\nu)\gamma}}\right\} \\
&\quad + e^{-k/16} + e^{-\frac{(\log\beta - j^*\mu^*)^2}{2j^*r}} \\
&\leq \sum_{\theta: a^*_\theta \neq a^*_{\theta^*}} \left\{\frac{e^{-\frac{k\epsilon^2\Delta_\theta^2}{4\gamma}}}{\left(1 - e^{-\frac{\epsilon^2\Delta_\theta^2}{2\gamma}}\right)} + 1 \wedge \sum_{t=\nu+k/2}^{\nu+k}\left(\mathbf{1}\left\{\kappa_{t-\nu}(\theta)^c\right\} + \mathbf{1}\left\{\kappa_{t-\nu}(\theta)\right\} e^{-\frac{\left(\epsilon(t-\nu)\Delta_\theta - Q_\nu^{(1)}(\theta)\right)^2}{2(t-\nu)\gamma}}\right)\right\} \\
&\quad + e^{-k/16} + e^{-\frac{\left(\log\beta - \frac{k(1-\epsilon)\mu^*}{4}\right)^2}{\frac{k}{2}(1-\epsilon)r}}, \qquad (15)
\end{aligned}$$

where we have denoted $1 \wedge x := \min\{x, 1\}$, and we have taken the minimum of the inner sum (over t) with 1 because probabilities are always bounded by 1.

The inner sum above for a fixed θ, clamped to 1, can be bounded as follows. Let $s_0 := \frac{2Q_\nu^{(1)}(\theta)}{\epsilon\Delta_\theta}$, so that $\epsilon s\Delta_\theta - Q_\nu^{(1)}(\theta) \geq \frac{\epsilon s\Delta_\theta}{2}$ whenever $s \geq s_0$. So $\forall\theta$, denoting $\gamma := r + 2D_{\max}^2$, and $1 \wedge x := \min\{x, 1\}$

we have

$$
\begin{aligned}
&1 \wedge \sum_{t=\nu+k/2}^{\nu+k} \left(\mathbf{1}\left\{\kappa_{t-\nu}(\theta)^c\right\} + \mathbf{1}\left\{\kappa_{t-\nu}(\theta)\right\} e^{-\frac{\left(\epsilon(t-\nu)\Delta_\theta - Q_\nu^{(1)}(\theta)\right)^2}{2(t-\nu)\gamma}} \right) \\
&= 1 \wedge \sum_{s=k/2}^{k} \left(\mathbf{1}\left\{\kappa_{s}(\theta)^c\right\} + \mathbf{1}\left\{\kappa_{s}(\theta)\right\} e^{-\frac{\left(\epsilon s\Delta_\theta - Q_\nu^{(1)}(\theta)\right)^2}{2s\gamma}} \right) \\
&\leq \mathbf{1}\left\{k/2 \geq s_0\right\} \sum_{s=k/2}^{k} e^{-\frac{\epsilon^2 s \Delta_\theta^2}{8\gamma}} + \mathbf{1}\left\{k/2 < s_0\right\} \cdot 1 \\
&\leq \mathbf{1}\left\{k \geq \frac{4Q_\nu^{(1)}(\theta)}{\epsilon\Delta_\theta}\right\} \left(\frac{e^{-\frac{k\epsilon^2\Delta_\theta^2}{16\gamma}}}{1 - e^{-\frac{\epsilon^2\Delta_\theta^2}{8\gamma}}} \right) + \mathbf{1}\left\{k < \frac{4Q_\nu^{(1)}(\theta)}{\epsilon\Delta_\theta}\right\} \cdot 1. \qquad (16)
\end{aligned}
$$

We are now in a position to obtain a bound on the (conditional) expected excess detection delay $\mathbb{E}^{(\nu,\theta^*)}\left[(\tau-\nu)^+\right]$ by integrating the tail and using (15) and (16):

$$
\begin{aligned}
&\mathbb{E}_Q^{(\nu,\theta^*)}\left[(\tau-\nu)^+\right] = \sum_{k=1}^{\infty} \mathbb{P}_Q^{(\nu,\theta^*)}\left[(\tau-\nu)^+ \geq k\right] = \sum_{k=1}^{\infty} \mathbb{P}_Q^{(\nu,\theta^*)}\left[\tau \geq \nu + k\right] \\
&\leq 20 + \frac{4\log\beta}{\mu^*(1-\epsilon)} + \sum_{k=\left\lceil \frac{4\log\beta}{\mu^*(1-\epsilon)} \right\rceil}^{\infty} e^{-\frac{\left(\log\beta - \frac{k(1-\epsilon)\mu^*}{4}\right)^2}{\frac{k}{2}(1-\epsilon)r}} \\
&+ \sum_{\theta: a_\theta^* \neq a_{\theta^*}^*} \left\{ \sum_{k=\left\lceil \frac{4\log\beta}{\mu^*(1-\epsilon)} \right\rceil}^{\infty} \frac{e^{-\frac{k\epsilon^2\Delta_\theta^2}{4\gamma}}}{\left(1 - e^{-\frac{\epsilon^2\Delta_\theta^2}{2\gamma}}\right)} + \frac{4Q_\nu^{(1)}(\theta)}{\epsilon\Delta_\theta} + \sum_{k=\left\lceil \frac{4\log\beta}{\mu^*(1-\epsilon)} \right\rceil}^{\infty} \frac{e^{-\frac{k\epsilon^2\Delta_\theta^2}{16\gamma}}}{\left(1 - e^{-\frac{\epsilon^2\Delta_\theta^2}{8\gamma}}\right)} \right\}.
\end{aligned}
$$

The third term above admits the bound

$$
\begin{aligned}
&\sum_{k=\left\lceil \frac{4\log\beta}{\mu^*(1-\epsilon)} \right\rceil}^{\infty} e^{-\frac{\left(\log\beta - \frac{k(1-\epsilon)\mu^*}{4}\right)^2}{\frac{k}{2}(1-\epsilon)r}} = \sum_{k=\left\lceil \frac{4\log\beta}{\mu^*(1-\epsilon)} \right\rceil}^{\left\lceil \frac{8\log\beta}{\mu^*(1-\epsilon)} \right\rceil} e^{-\frac{\left(\log\beta - \frac{k(1-\epsilon)\mu^*}{4}\right)^2}{\frac{k}{2}(1-\epsilon)r}} + \sum_{k=\left\lceil \frac{8\log\beta}{\mu^*(1-\epsilon)} \right\rceil}^{\infty} e^{-\frac{\left(\log\beta - \frac{k(1-\epsilon)\mu^*}{4}\right)^2}{\frac{k}{2}(1-\epsilon)r}} \\
&\leq 1 + \frac{4\log\beta}{\mu^*(1-\epsilon)} + \frac{e^{-\frac{\mu^*\log\beta}{4r}}}{1 - e^{-\frac{(\mu^*)^2(1-\epsilon)}{32r}}},
\end{aligned}
$$

while each summand corresponding to θ in the final term is bounded as

$$
\sum_{k=\left\lceil \frac{4\log\beta}{\mu^*(1-\epsilon)} \right\rceil}^{\infty} \frac{e^{-\frac{k\epsilon^2\Delta_\theta^2}{4\gamma}}}{\left(1 - e^{-\frac{\epsilon^2\Delta_\theta^2}{2\gamma}}\right)} + \frac{4Q_\nu^{(1)}(\theta)}{\epsilon\Delta_\theta} + \sum_{k=\left\lceil \frac{4\log\beta}{\mu^*(1-\epsilon)} \right\rceil}^{\infty} \frac{e^{-\frac{k\epsilon^2\Delta_\theta^2}{16\gamma}}}{\left(1 - e^{-\frac{\epsilon^2\Delta_\theta^2}{8\gamma}}\right)} \leq \frac{4Q_\nu^{(1)}(\theta)}{\epsilon\Delta_\theta} + \frac{2e^{-\frac{\epsilon^2\Delta_\theta^2\log\beta}{4\gamma\mu^*(1-\epsilon)}}}{1 - e^{-\frac{\epsilon^2\Delta_\theta^2}{8\gamma}}},
$$

giving

$$
\mathbb{E}_Q^{(\nu,\theta^*)}\left[(\tau-\nu)^+\right] \leq 21 + \frac{8\log\beta}{\mu^*(1-\epsilon)} + \sum_{\theta: a_\theta^* \neq a_{\theta^*}^*} \left(\frac{4Q_\nu^{(1)}(\theta)}{\epsilon\Delta_\theta} + \frac{2e^{-\frac{\epsilon^2\Delta_\theta^2\log\beta}{4\gamma\mu^*(1-\epsilon)}}}{1 - e^{-\frac{\epsilon^2\Delta_\theta^2}{8\gamma}}} \right),
$$

where (recall) $\mu^* := D\left(\theta^*(a_{\theta^*}^*)||\theta_0(a_{\theta^*}^*)\right)$ and $\gamma = r + 2D_{\max}^2$.

Taking expectation under the $\mathbb{P}^{(\infty)}$ distribution of Q_ν completes the proof of Theorem 3.

Lemma 4. *For any sequence $x_1, \ldots, x_{t-1}$ and $\nu \in [t]$,*

$$
\max_{i=1}^{t} \sum_{\ell=i}^{t-1} x_\ell \leq \left(\max_{i=1}^{\nu-1} \sum_{\ell=i}^{\nu-1} x_\ell \right)^+ + \max_{i=\nu}^{t} \sum_{\ell=i}^{t-1} x_\ell = \max_{i=1}^{\nu} \sum_{\ell=i}^{\nu-1} x_\ell + \max_{i=\nu}^{t} \sum_{\ell=i}^{t-1} x_\ell.
$$

Proof of Lemma 4. The lemma is a consequence of the elementary result that $\max_i(a_i + b_i) \leq \max_i a_i + \max_i b_i$ for any discrete collection of numbers $\{a_i\}_i$, $\{b_i\}_i$. □

C Experiment Details

This section describes in detail the setup and methodology followed for obtaining the results in Section 6. It also includes additional results, both for the synthetic and audio sensing settings, that explore the impact of various problem parameters on performance.

C.1 Version of the ϵ-GCD algorithm used in experiments

In all our experiments, we used the *full data-MLE* implementation of the ϵ-GCD template, as given in Algorithm 1. The only difference of this algorithm from the *exploration data-MLE* version given in the main text (Algorithm 1) is that the estimate $\hat{\theta}_t$ for the post-change distribution is computed using data from *all* previous rounds, regardless of exploration or exploitation.

Algorithm 1 ϵ-GCD (Full data-MLE version)

Input: Exploration rate $\epsilon \in [0, 1]$, Pre-change parameter θ_0, Post-change parameter set $\Theta^{(1)}$, Stopping threshold $\beta > 0$, Exploration distribution over actions π, Observation function $g_\theta(x, a)$ $\forall(\theta, x, a) \in \Theta^{(1)} \times \mathcal{X} \times \mathcal{A}$.
Init: $Q_1^{(1)}(\theta) \leftarrow 0$, $Q_1^{(2)}(\theta) \leftarrow 0\ \forall\theta \in \Theta^{(1)}$ {CUSUM statistics based on overall and exploitation-only data, per parameter}
for round $t = 1, 2, 3, \ldots$ **do**
  if $\max_{\theta\in\Theta^{(1)}} Q_t^{(2)}(\theta) \geq \beta$ **then**
    break {Stop sampling and exit}
  end if
  Sample $E_t \sim \text{Ber}(\epsilon)$ independently
  if $E_t == 1$ **then**
    Play action $A_t \sim \pi$ independently {Explore}
    Get observation X_t
  else
    Compute $\hat{\theta}_t = \text{argmax}_{\theta\in\Theta^{(1)}} Q_t^{(1)}(\theta)$ {Most likely post-change distribution based on all past data}
    Play action $A_t = \text{argmax}_{a\in\mathcal{A}} D\left(\theta^{(0)}(a)||\hat{\theta}_t(a)\right)$
    Get observation X_t
    Set $\forall\theta \in \Theta^{(1)} : Q_{t+1}^{(2)}(\theta) \leftarrow \left(Q_t^{(2)}(\theta) + g_\theta(X_t, A_t)\right)^+$ {Update exploitation CUSUM statistics}
  end if
  Set $\forall\theta \in \Theta^{(1)} : Q_{t+1}^{(1)}(\theta) \leftarrow \left(Q_t^{(1)}(\theta) + g_\theta(X_t, A_t)\right)^+$ {Update overall data CUSUM statistics}
end for

The chief reason to prefer the exploration-only version in the main text is that the theoretical analysis of its detection delay is slightly simpler than the full-data version. This is because the exploration-only CUSUM statistic (queue) changes from one time to the next in an essentially memoryless manner since the sensing action is chosen independent of the past.

On the other hand, we preferred the full data-version in experiments since it was noticed to offer slightly better numerical performance. We remark that the full-data MLE version can also be analyzed rigorously for its detection delay[9] in a manner similar to that of Algorithm 1, with an essentially similar delay guarantee. However, one will have to content with an extra overhead in the additive term of the detection delay (i.e., the last term of (1)), due to not being able to apply time-uniform maximal inequalities for martingales (see (7), (8) in Section B.2) but resort to a slightly worse union bound over time.

C.2 Algorithms compared in the experiments

We compare the performance of the following adaptive sensing change detection algorithms in all our experimental settings:

[9]The false alarm rate analysis remains the same as there is no change to the stopping criterion in the algorithm.

1. The ϵ-GCD algorithm ('EG' in plots), used with $\epsilon = 0.2$
2. Uniform Random Sampling ('URS'): The non-adaptive sensing rule that plays an action drawn uniformly at random from a given set of actions; this is ϵ-GCD with $\epsilon = 0$
3. Oracle Sampling ('Oracle'): The sensing rule that always plays the most informative action knowing the post-change distribution in advance: $\mathrm{argmax}_{a\in\mathcal{A}}\, D\left(\theta^*(a)||\theta^{(0)}(a)\right)$.

C.3 Synthetic experiments

We conduct change detection experiments on a line graph, of size N, serving as an ambient space. Nodes of the graph are interpreted as physical locations, and take values in $[N]$. Nodes j, k are connected if $|j-k| = 1$. Each node $n \in [N]$ offers a Gaussian-distributed observation depending on the changepoint $\nu \in \mathbb{N}$. In particular, the signal (observations from all nodes) at time t is a random vector $S(t) = (S_n(t))_{n\in[N]} \in \mathbb{R}^N$, where

$$S_n(t) = \theta_n \mathbf{1}\{t \geq \nu\} + W_n(t),\ t = 0, 1, 2, \ldots \tag{17}$$

$\theta := (\theta_n)_{n\in[N]}$ represents the post-change parameter and $W_n(t) \in \mathcal{N}(0, \sigma^2)$ are IID Gaussian random variables for $n \in [N]$ and across time $t = 0, 1, \ldots,$, and represents observation noise. We also choose $\sigma^2 = 1/2$ for all our synthetic experiments. Note that in essence, this setup has the pre-change parameter set to zero ($\theta = 0 \in \mathbb{R}^N$).

Isolated and Structured Anomalies. We consider two types of the vector of change parameters:
(a) Isolated singleton change, namely, $\theta_n \in \{0,\ 1\}$ and $\sum_{n\in[N]} \theta_n = 1$;
(b) Structured K-change: We consider parameter changes with[10] $|\text{Supp}(\theta)| = K$, and the nodes (components) corresponding to the non-zero support are connected. As such the collection of anomalies is $N - K + 1$ corresponding to different starting positions.

Diffuse and 'Pointy' Action sets. In a parallel fashion we allow actions to be vertices of the N-hypercube, $a_n \in \mathcal{A} \subset \{0,\ 1\}^N$, and the action sets to be either *pointy* ($\mathcal{A}_1$), namely, $|\text{Supp}(a)| = 1$ $\forall a \in \mathcal{A}_1$, which allows probing only single nodes, or *diffuse* ($\mathcal{A}_2$), where only a connected subset of nodes can be queried. In either case, the observation received on an action, $a \in \mathcal{A}$ is given by $X_a = \langle \frac{a}{\|a\|_2}, S\rangle$, where we impose the normalization because we want to maintain the same signal-to-noise ratio (SNR) across different types of probes.

We reported results for pointy actions and isolated anomalies in Sec. 6. We will describe experiments with other scenarios here.

Structured Anomalies and Diffuse Action Sets.
We experiment with Structured K-change as described above with $K = 5$. The action sets are diffuse:

$$\mathcal{A} = \{a \in \{0,1\}^N : \|a\|_2 = 1, |\text{Supp}(a)| = 5,\ a \text{ is connected.}\}$$

This means that corresponding to each anomalous change, there is an action (unknown to the learner) that perfectly overlaps with the entire anomalous change. Furthermore, there are several other actions that partially overlap with the structured anomaly. As a result there is a higher probability of detecting anomalies.

As in Sec. 6 we tabulate results for change point $\nu = 40$, for different sizes of graph. Our results are based on 5000 Monte Carlo runs. The mean and standard deviation for change point $\nu = 40$ is reported in Table 2. We see that, although diffuse, such actions appear to improve detection delay for ϵ-GCD in comparison to the case considered in Sec. 6. This is to be expected because the number of structured anomalies are smaller. For instance for a graph of size 10, we only have 5 anomalous parameter changes. Furthermore, we can detect anomalies even when the actions only partially overlap with the anomaly. Thus change detection methods now have a larger probability to detect parameter changes in contrast to isolated anomalies.

We also observe that ϵ-GCD is still as effective, and closely mirrors Oracle performance. We do not tabulate the effect of different changepoints here. This is because, we notice that the changepoint parameter ν has no noticeable effect when graph size is held constant for all of the reported methods (Oracle, URS, and ϵ-GCD).

[10] We define $\text{Supp}(x) := \{i : x_i \neq 0\}$.

Size	Oracle	ϵ-GCD(Full)	ϵ-GCD(Theory)	URS
10	51 ± 7	64 ± 12	90 ± 21	96 ± 14
15	51 ± 7	71 ± 13	100 ± 27	163 ± 26
20	51 ± 7	74 ± 14	113 ± 35	237 ± 40
25	51 ± 7	80 ± 17	126 ± 46	310 ± 52

Table 2: Structured Anomalies with Diffuse Actions: Observed mean and standard deviation for the simulated stopped time for varying graph-sizes with diffused action set for change occurring at $\nu = 40$.

Structured Anomalies with Pointy Action Sets.
Here we experiment with $K = 5$ as in the setup above but examine the effect of pointy action sets. As a result our actions can only probe some component of the anomaly. Observe that the anomaly is spread across a larger region. Therefore, a pointy anomaly can only sample a small part of the parameter change in any round.

We report mean and variance for expected delay for change point $\nu = 40$, for different graph sizes over 5000 Monte Carlo runs in Table 3.

Size	Oracle	ϵ-GCD(Full)	ϵ-GCD(Theory)	URS
10	150 ± 27	192 ± 41	289 ± 124	299 ± 57
15	149 ± 27	210 ± 49	340 ± 175	448 ± 85
20	150 ± 27	204 ± 44	350 ± 213	597 ± 118
25	149 ± 27	213 ± 48	395 ± 246	746 ± 145

Table 3: Structured Anomalies with Pointy Action Sets. Observed mean and standard deviation for the simulated stopped time for varying graph-sizes at $\nu = 40$.

In this experiment both Oracle and ϵ-GCD exhibit larger delays. The reason now is that pointy anomalies can only sample a single component, and as such a component in the anomalous region exhibits smaller change, and so it takes a longer time to detect. Again, no noticeable impact of varying changepoint on delay was observed.

Isolated Anomalies with Diffuse Action Sets.
Here we consider the case where the anomalies are isolated but the action sets are diffuse. Our results (mean and variance) over 5000 Monte Carlo runs for changepoint $\nu = 40$ is tabulated in Table 4.

Size	Oracle	ϵ-GCD+	ϵ-GCD(Theory)	URS
10	249 ± 35	277 ± 40	321 ± 49	497 ± 74
15	250 ± 35	281 ± 40	326 ± 54	549 ± 81
20	249 ± 35	297 ± 42	370 ± 84	998 ± 151
25	249 ± 35	324 ± 52	509 ± 237	1044 ± 157

Table 4: Isolated Anomalies with Diffuse Action Sets. Observed mean and standard deviation for the simulated stopped time for varying graph-sizes for change occurring at $\nu = 40$.

Among all of the different scenarios, this setup has uniformly larger expected delay across all of the methods (Oracle, ϵ-GCD and URS). This is not surprising considering the fact that isolated anomalies when probed with diffuse actions manifest as substantially smaller change. This is because a diffuse action, spread across 5 locations, is capable of collecting only a 5th of the energy of the anomaly.

D Experiment Details for Real-World Audio Dataset

Recall from Sec. 6 we explored changepoint detection on the MMII dataset. We pointed out that we used reconstruction errors of auto-encoders, and modeled these errors with Gaussians. Here we provide more details and additional experiments on the dataset.

Audio Processing. For each audio-stream we train auto-encoders on normal data using mel-spectrogram features. We train different autoencoders for different machine ids. We use mean of reconstruction errors from each of these machines when there are no anomalies, and construct the pre-change parameter vector. Similarly, we use mean of reconstruction errors from each of these machines when there are anomalies and construct post-change parameters.

To compute reconstruction errors, we adopt the autoencoder architecture as used in Section 4 in [Pur+19b]. We also make use of publicly available code to train autoencoders with mel-spectrogram features of normal data as inputs. We use the same parameters that are used in [Pur+19b], to extract mel-spectrogram features from a given audio input. We assume that, for a given audio stream, there is only one anomaly, and that anomaly is present in only one of the machine ids.

The resulting pre-change parameters across the 4 machines and the post-change parameters under an anomaly are displayed below:

Machine ID	**Mean reconstruction error**	
	Normal	**Abnormal**
00	7.816003	18.043417
02	7.728631	12.879204
04	12.029381	15.425252
06	9.34813	10.788003

Table 5: Mean reconstruction errors for machine ID $00, 02, 04, 06$ under normal and abnormal operation

Using the notation of our synthetic experiment, our setup here can be described as the case with isolated anomalies (i.e., only one machine has an anomaly), and pointy actions (i.e., we can only query one machine at any time).

Experiment. In addition to the $\nu = 6$ case, which we reported in the main paper, we simulate changepoints for $\nu = 21$. This corresponds to 210 seconds. We do this by introducing anomalies in machine bearing ID00 as follows. We concatenated 21 normal files and 39 abnormal files chosen uniformly at random from machine ID00. For the other machines we concatenated 60 normal files at random. The 60 files correspond to 600 seconds. Our changepoint corresponds to 21st file, which we denote as $\nu = 21$ and our task is to detect this change. Note that both the machine ID and the changepoint is not known to the learner. Our results are depicted as histograms for changepoints of anomaly detection in Fig. 4.

As observed Oracle method has a small variance, and the histogram is concentrated at around 25, which is about 40 seconds delay. ϵ-GCD also exhibits small delay, but its variance is somewhat larger in this context. URS expected delay and variance are substantially larger. This demostrates the gains due to adaptive processing.

E Detailed Survey of Related Work

?? Changepoint detection deals with identifying points in time when probability laws governing a stochastic process changes abruptly. The problem of changepoint detection has been widely studied, dating back to the pioneering works of Page [Pag54] and Lorden [Lor71].

Online change detection focuses on situations where the data is obtained incrementally over time, and one must infer whether a change has occurred at each time. A large part of the online change detection literature, like our paper, adopts a frequentist approach, and, in particular, utilizes parametric models for pre-change and post-change distributions. In this context, the CUSUM algorithm [Pag54] and its variants such as the generalized likelihood ratio statistics have been proposed, in which a change is announced when the likelihood ratio statistic exceeds a threshold. While there are a number of prior works on this topic (see [BN93; CG12; VB14]), specific attention to finite time (i.e., non-asymptotic) guarantees on detection delay is more recent [LX10; Mai19], and as such remains somewhat open

Figure 3: Histogram of reconstruction errors under normal and abnormal operation of machine id $00, 02, 04, 06$.

([AC17]). While changepoint detection has also been studied from a Bayesian perspective, much of this literature has focused on the batch setting (see [Gun+21; AM07; FL07]). Though recent works have begun to focus on Bayesian online detection, there has been little work (apart from [AMF20]) on proving finite time guarantees.

Our work is motivated by costs imposed on data collection due to resource constraints. In this context, while adopting a frequentist perspective, we propose methods for adaptive online data selection for multi-stream time-series data. Recent works have begun to focus attention on adaptive online data collection for changepoint detection from both frequentist [ZM20] and Bayesian [OGR10; Gun+21] perspectives. Additionally, while different from our focus, we mention in passing that methods for active change-point detection [HKK19], where the task is to adaptively determine the next input have also been proposed. Furthermore, there are a number of works that focus on bandit regret minimization for non-stationary time-series data [GM11; LLS18; MS13]. While these works are related we note that regret minimization is a fundamentally different objective from changepoint identification where the goal is to minimize detection delay.

We will outline similarities and differences between our work and closely related prior works. From a practical perspective, [OGR10] is similar to our work in that they too motivate their approach from a sensor network viewpoint, where sensors may undergo faults or changepoints exhibited due to environmental factors. They propose a Bayesian formalism largely based on the well-known Bayesian online change detection (BOCD) method [AM07], and leverage Gaussian processes (GP) for modeling. The GP perspective allows for tractable sequential time-series prediction and sensor selection. Recently, [Gun+21] have proposed to extend the BOCD approach to incorporate costs in making decisions for real-time data acquisition in multi-fidelity sensing scenarios. Different from these perspectives, our approach is frequentist, and does not impose action-specific costs and distributional constraints on the underlying latent parameters or on the changepoints. Furthermore, we derive finite time performance bounds, which is not a focus of any of these works.

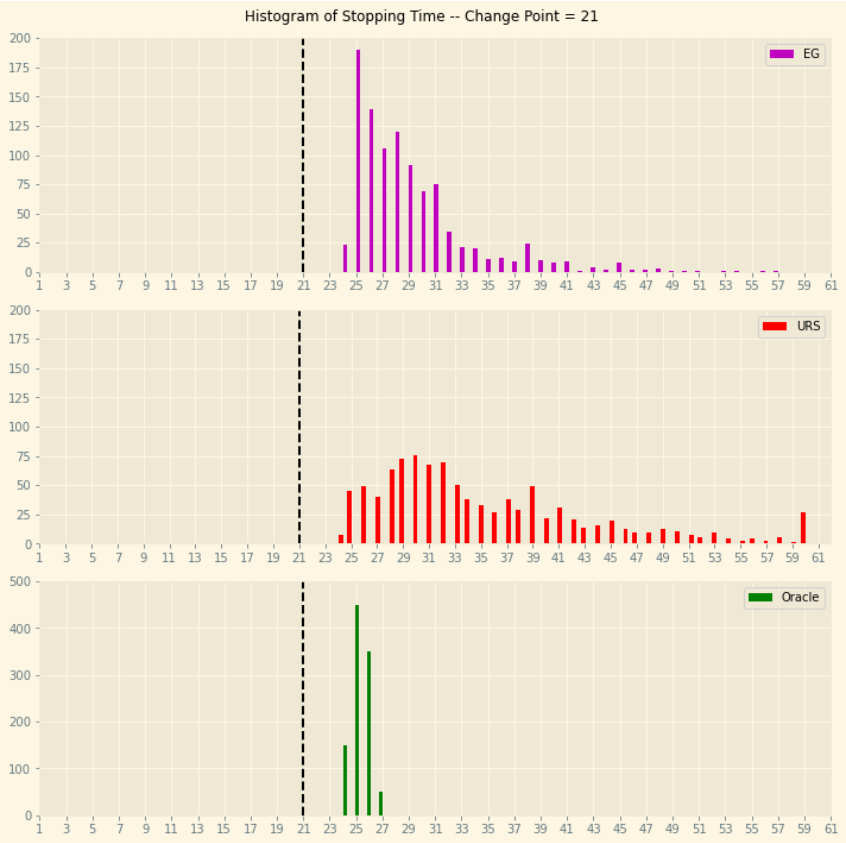

Figure 4: Audio based change detection of machine anomaly: Histogram of stopping times by URS, Oracle, and ϵ-GCD (EG) for changepoint $\nu = 21$.

Similar to our work, [ZM20] also adopts a frequentist perspective and proposes a method for bandit changepoint detection based on the well-known Shiryaev-Roberts-Pollak scheme. To adaptively choose data streams/sensors, they utilize Thompson sampling [Tho33] to balance exploration of different data streams for acquiring knowledge with exploitation of informative streams. In addition, they present theoretical properties, and show that their method does not trigger false alarm too soon. However, detection delays under false alarm constraints are not explicitly characterized, which is a key challenge.

In summary, the principal difference between these prior works on bandit online changepoint detection and our ϵ-GCD is that we are able to explicitly characterize information theoretic lower bounds on expected detection delay under a false alarm constraint. Furthermore ϵ-GCD is natural variant of CUSUM, and our explicit analysis shows that it exhibits finite-time performance guarantees on expected detection delay and matches our lower bound at low false alarm rates. This leads to optimality that is absent in prior work.

F Discussion and Future Work

This work has laid down a principled approach to exploration with information-limited sensing to rapidly detect changes in distribution. Specifically, we have shown that relatively 'simple' (epsilon-greedy) forced exploration is sufficient to obtain detection delays comparable to an oracle who knows the post-change distribution beforehand.

As such, this study represents only an initial attempt to understand the limits of adaptive sensing for change detection, and opens up a host of interesting avenues for further investigation. These include (a) the possibility of 'more adaptive' exploration approaches, such as confidence-set or posterior sampling-based methods, that could improve the delay for learning a good guess of the post-change distribution (the second term in the detection delay bound), (b) adaptive sensing when

both the pre change and post change distribution is unknown, which also entails learning the default distribution online, (c) extensions to continuous parameter spaces, (d) detecting multiple changes that occur continually over time, and (e) studying the adaptive change detection problem for Markovian dynamics or controlled processes.