On Low-Complexity Quickest Intervention of Mutated Diffusion Processes Through Local Approximation

Qining Zhang
qiningz@umich.edu
University of Michigan, Ann Arbor
Ann Arbor, Michigan, USA

Honghao Wei
honghaow@umich.edu
University of Michigan, Ann Arbor
Ann Arbor, Michigan, USA

Weina Wang
weinaw@cs.cmu.edu
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA

Lei Ying
leiying@umich.edu
University of Michigan, Ann Arbor
Ann Arbor, Michigan, USA

ABSTRACT

We consider the problem of controlling a mutated diffusion process with an unknown mutation time. The problem is formulated as the quickest intervention problem with the mutation modeled by a change-point, which is a generalization of the quickest change-point detection (QCD). Our goal is to intervene in the mutated process as soon as possible while maintaining a low intervention cost with optimally chosen intervention actions. This model and the proposed algorithms can be applied to pandemic prevention (such as Covid-19) or misinformation containment. We formulate the problem as a partially observed Markov decision process (POMDP) and convert it to an MDP through the belief state of the change-point. We first propose a grid-approximation approach to calculate the optimal intervention policy, whose computational complexity could be very high when the number of grids is large. In order to reduce the computational complexity, we further propose a low-complexity threshold-based policy through the analysis of the first-order approximation of the value functions in the *"local intervention"* regime. Simulation results show the low-complexity algorithm has a similar performance as the grid-approximation approach and both perform much better than the QCD-based algorithms.

CCS CONCEPTS

• **Mathematics of computing** → **Markov processes**; • **Networks** → *Network performance analysis*.

KEYWORDS

Quickest intervention, change-point, POMDP, local intervention

ACM Reference Format:

1 INTRODUCTION

Mutated diffusion processes such as variants of SARS-Cov-2 (Covid-19), malicious behaviors in communication networks, and manipulated information on online social networks can create serious public health issues and severe financial crises, with catastrophic social and economic consequences. These highly consequential information and virus "mutations" often occur in subtle and sometimes random ways, which makes it difficult to control/intervene at their early stages. For example, the Delta variant of Covid-19 has quickly become the dominant strain in many countries since first detected in March 2021; malicious devices may merely deviate subtly from allocated protocols at the early stage; and misinformation is often embedded in real news by manipulating only a few but not all details of a true news story. The focus of this paper is to develop both theories and algorithms to understand *when and how to intervene in a diffusion process, which may have mutated, to control its damage?*

In this paper, we consider a model where an agent receives a sequence of observations assumed to be drawn from a distribution. Here, an observation may represent the symptom of a Covid-19 patient or whether an online platform user has retweeted the news after reading it on social media. A mutation that occurs at a time unknown to the agent changes the observation distribution, while the sample space remains the same. At each time t, given the sequence of observations, the agent decides whether to intervene to control the mutated diffusion process. If so, how strict the intervention action should be to limit its impact, e.g., whether to suggest or mandate mask-wearing, and whether to ask voluntary quarantine or impose a strict lockdown.

We call this problem "quickest intervention" because it is closely related to quickest change detection (QCD), where the goal is to quickly detect a change-point based on sequential observations. The QCD problem was first proposed and studied by Shiryev [14], where he studied a parameterized observation distribution with a geometric prior for the change-point and proposed the well-known Shiryev's test. Lorden extended Shiryev's theory with the mini-max criterion by replacing the geometric prior with any unknown non-random change-point [7]. He also showed the CuSum test

from [11] is asymptotically optimal as false alarm requirements become stricter, while the exact optimality was proved by Moustakides [10] and Ritov [13]. Improvements based on CuSum have been proposed in various works including [12]. Our work is closely related to the Bayesian regime summarized in Shiryev's book [15]. However, QCD focuses only on detecting the change-point and does not consider taking actions that will change the distribution of the observations. A recent paper [1] did consider intervention actions in QCD, but the problem was studied mostly via simulations. In terms of applications, [18] studied the quickest detection of misinformation using the optimal stopping theory, which again did not consider the intervention.

Since an action changes the underlying diffusion process in "quickest intervention", there are fundamental differences between our problem and pure detection. The problem in fact falls into the category of partially observed Markov decision process (POMDP) [4], where an agent makes decisions while receiving sequential observations and inferring unobserved states (change-point or mutation). It is well-known that solving general POMDPs is extremely difficult. Only limited structural results and low-complexity solutions with provable performance for general POMDPs were studied. For example, [8] established the monotonicity of value functions under the monotone likelihood ratio (MLR) ordering condition. [5] later improved the result by developing a low-complexity myopic policy. However, in general, even establishing sufficient conditions that the optimal policy of a POMDP possesses a threshold structure is extremely hard. In this paper, we first formulate the quickest intervention problem as a POMDP and then develop low-complexity algorithms and theoretical results in the "local intervention" regime.

The main contributions of this paper are summarized below:

(1) Problem Formulation: We formulate the quickest intervention problem as a discrete-time process with an unknown change-point (mutation) and multiple intervention actions that can be used to control the diffusion process. The objective of the problem is to identify a policy that minimizes the total propagation and intervention costs. This formulation is analogous to QCD but involves decision-making that affects the underlying diffusion process.

(2) POMDP Approach: The problem is actually a POMDP where the hidden state is the change-point. We convert it into a fully observed MDP via belief states. Grid approximation is then proposed to compute the optimal policy numerically.

(3) "Local Intervention" Regime: To overcome the computational complexity of grid approximation when the number of grids is large, the key contribution of this paper is to develop a low-complexity algorithm based on the first-order approximation of value functions in the "local intervention" regime. With a derived approximated Bellman equation, we prove that the action-value function in the "local intervention" regime is nearly submodular, and the policy that solves the approximated Bellman equation has a threshold structure. Furthermore, we derive upper bounds on the thresholds with closed-form expressions. Based on the upper bounds, we propose a low-complexity algorithm that increases the intervention levels when the belief exceeds the corresponding threshold. Interestingly, our numerical results show that the low-complexity algorithm is near-optimal even beyond the "local intervention" regime, which demonstrates its effectiveness (near-optimal performance) and efficiency (low computational complexity).

2 MOTIVATING EXAMPLES

In this section, we provide motivating examples in the application domain that fall into our problem formulation.

2.1 Covid-19 Pandemic Prevention

As the Covid-19 disease spreads all around the world, prevention and detection policies are implemented in different countries to relieve the global situation. Specifically, consider a medical institution that takes care of Covid-19 patients. The manager of this institution is receiving observations representing the severeness of symptoms from the patients each day. According to these symptoms, the manager decides whether a more harmful new variant has infected this institution and whether to take actions such as using heavier masks and implementing quarantines to stop the spread of new variants.

2.2 Misinformation Intervention on Social Networks

On social networks such as Twitter and Weibo, millions of fake news posts are created every day by social bots, which affect the judgement of real users (people). Consider an administrator who detects misinformation on its platform. It tries to set up different levels of labels and warnings for potential fake news. Once a piece of information is retweeted, the administrator receives a new observation on the features of the retweeting user. Based on these features, the administrator decides whether the potential misinformation has contaminated real user communities, and which intervention action to implement to prevent users from retweeting it.

2.3 Anomaly Detection in Communication Networks

A communication link is provided and maintained by a service provider to multiple users or devices for transmitting data packets with a designated distribution of rates. Some user may become malicious at certain points and start to send packets with an higher rate for a long period time. Under such circumstances, the service provider may want to detect such behaviors and then take multiple interventions to throttle the data rate for some users.

3 MODEL

In this section, we introduce the notations, the diffusion and intervention models with a change-point representing the mutation, and the problem formulation.

3.1 Notation

We use $1_{\{\cdot\}}$ to denote the indicator function, Pr to denote the probability measure, and $\mathbb{E}_X$ to denote expectation with respect to random variable X. If the subscript is omitted, the expectation is taken with respect to all randomness. Vectors and matrices are written in boldface $\boldsymbol{a}$ and $\boldsymbol{A}$, with a_i or $a(i)$ indicating the i-th entry of $\boldsymbol{a}$ and $A_{i,j}$ indicating the (i, j)-th element of $\boldsymbol{A}$.

3.2 Sequential Diffusion Model

We consider a discrete-time process as shown in Fig. 1 with a change-point τ. An agent is observing this process and needs to intervene after a harmful mutation occurs at the change-point τ. Let $t \in$
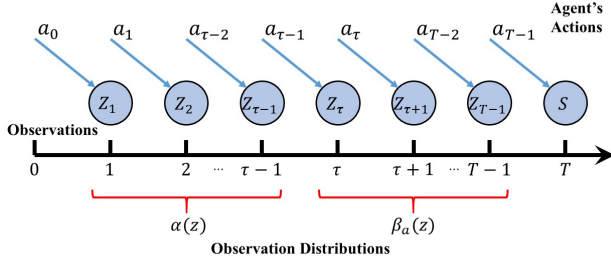
Figure 1: Illustration of the sequential diffusion model. At each time step, the agent receives an observation z_t (blue circle), and then it makes an intervention a_t to affect the observation z_{t+1} in the next time step.

$\{1, 2, \cdots, T\}$ be the index of time slots, where the time horizon T is a random variable following a geometric distribution with parameter $1 - \rho$, i.e.,

$$\Pr(T = k) = \rho^{k-1}(1 - \rho). \tag{1}$$

We assume the agent does not know T a priori, but is aware of whether the process has terminated or not (e.g. a cure has been discovered or the disease suddenly disappears). For simplicity, we use $s_t = 1_{t \geq T} \in \{0, 1\}$ to indicate whether the process has ended, i.e., $s_t = 1$ means it has ended and $s_t = 0$ means otherwise.

Let $\tau \in \{1, 2, \cdots, T\}$ denote the change-point which is also unknown to the agent throughout the process. The change-point τ divides the time horizon into two phases: the pre-change phase before τ and the post-change phase after τ. We assume the change-point τ follows a geometric prior distribution with parameter λ:

$$\Pr(\tau = k) = \lambda(1 - \lambda)^{k-1}. \tag{2}$$

For simplicity, we also denote $\theta_t = 1_{t \geq \tau} \in \{0, 1\}$ to be the indicator whether the change point has happened.

At each time slot t, the agent receives an observation $z_t \in \{1, 2, \cdots, Z\}$, which will be used by the agent to determine whether the change-point has occurred and how to intervene. We assume in the absence of any intervention actions, z_t follows an i.i.d. distribution $\alpha(z)$ in the pre-change phase, and follows another i.i.d. distribution $\beta_0(z)$ in the post-change phase:

$$\Pr(z_t = z | \theta_t = 0) = \alpha(z), \tag{3a}$$
$$\Pr(z_t = z | \theta_t = 1) = \beta_0(z). \tag{3b}$$

We remark that in this paper, we consider a harmful mutation occurs at the change-point τ and turns the original benign process into a harmful one. The terminal time T represents the time at which either a cure is discovered or the mutation disappears. Considering the Covid-19 pandemic, the observations z_t could be viewed as the symptoms of the t-th patient. A higher value of z_t represents a severer symptom. For example, $z_t = 1$ represents an asymptomatic patient while $z_t = Z$ represents hospitalization.

3.3 Sequential Intervention Model

While receiving observations, the agent needs to decide whether to intervene and what actions to take. We assume the agent can take A different actions representing A different intervention levels, i.e., $a_t \in \{1, \cdots, A\}$, or it can choose to idle with $a_t = 0$. We further assume that the intervention is only effective on the already mutated diffusion process, i.e., when $\theta_t = 1$. The intervention will shift the original probability distribution $\beta_0(z)$ to a new distribution $\beta_a(z)$ according to the action a which the agent chooses. Therefore, depending on whether the mutation has occurred, the observation z_t is an i.i.d. sample from one of the following distributions:

$$\Pr(z_t | \theta_t = 0, a_{t-1} = a) = \alpha(z_t), \quad \forall a \in \{0, \cdots, A\}, \tag{4a}$$
$$\Pr(z_t | \theta_t = 1, a_{t-1} = a) = \beta_a(z_t), \quad \forall a \in \{0, \cdots, A\}. \tag{4b}$$

We remark that larger a_t values indicate stricter intervention levels. Considering controlling Covid-19 pandemic, $a_t = 1$ may represent voluntary social distancing and $a_t = A$ may mean strict lock-down. Throughout the paper, we assume that with the highest intervention level, we can suppress the mutated diffusion process to be as the original one, i.e., $\boldsymbol{\beta}_A = \boldsymbol{\alpha}$. Moreover, we assume that the agent can only gradually increase the intervention level a_t and cannot decrease it. In other words, the intervention cannot be relaxed until the diffusion ends. We further impose the condition that the agent can only increase the intervention level by one at each time step, i.e., $a_{t-1} \leq a_t \leq a_{t-1} + 1$. This is a technical assumption and does not affect the generality of our solution because we can choose the duration of each time slot to be sufficiently small so that the intervention level can go up multiple levels in a short period of time. We will show in simulations that leveraging this assumption doesn't harm our results empirically too much.

When controlling the process, the agent needs to consider the cost of an intervention action (e.g. the economic cost of lockdown) and the cost of the diffusion propagation (e.g. treatment of infected patients). We define c_p^z to be the propagation cost of observation z, which can be viewed as the treatment of a patient with symptom z, and define c_i^a to be the intervention cost of taking action a at any time t. We assume that c_i^a and c_p^z are monotonically non-decreasing with a and z respectively, since stricter actions require more resources and severer symptoms require more treatments. Especially, the intervention cost of idle equals zero, i.e., $c_i^0 = 0$.

3.4 Dominance Assumptions

In order to model the strictness of intervention actions, we assume the distribution vector $\boldsymbol{\beta}_a$ of observations z_t follows first order stochastic dominance and monotone likelihood ratio (MLR) ordering assumptions defined below.

Definition 3.1 (Stochastic Dominance). For two probability mass functions $\beta(z)$ and $\tilde{\beta}(z)$ on the same probability space, we say $\boldsymbol{\beta}$ stochastically dominates $\tilde{\boldsymbol{\beta}}$, denoted by $\boldsymbol{\beta} \geq_s \tilde{\boldsymbol{\beta}}$, if

$$\sum_{z=j}^{Z} \beta(z) \geq \sum_{z=j}^{Z} \tilde{\beta}(z), \quad \forall j \in \{1, 2, \cdots, Z\}. \tag{5}$$

Definition 3.2 (MLR Ordering). Given two probability mass functions $\beta(z)$ and $\tilde{\beta}(z)$ on the same probability space, we say $\boldsymbol{\beta}$ MLR

dominates $\tilde{\beta}$, denoted by $\beta \geq_r \tilde{\beta}$, if

$$\frac{\beta(z)}{\tilde{\beta}(z)} \leq \frac{\beta(z')}{\tilde{\beta}(z')} \quad \forall z \leq z'. \tag{6}$$

The two assumptions $\beta \geq_s \tilde{\beta}$ and $\beta \geq_r \tilde{\beta}$ both imply that under distribution β, it is more likely to receive a higher value observation. Moreover, MLR ordering $\beta \geq_r \tilde{\beta}$ is a stronger assumption which directly implies stochastic dominance $\beta \geq_s \tilde{\beta}$ [4]. In this paper, we assume the distribution of a less strict action dominates the distribution of a stricter one, i.e.,

$$\beta_a(z) \leq_r \beta_{a-1}(z), \quad \forall a \in \{1, 2, \cdots, A\}. \tag{7}$$

We also assume that in the post-change phase after time τ, stricter actions achieve lower overall cost:

$$\mathbb{E}[c_p^{z_t} + c_i^{a_t} | \theta_t = 1] < \mathbb{E}[c_p^{z_t} + c_i^{a_t-1} | \theta_t = 1], \quad \forall a_t \geq 1. \tag{8}$$

3.5 Problem Formulation

The objective of the agent is to minimize the overall cost including both the propagation and intervention costs, without knowing τ and T. It requires the agent to detect the change-point τ as soon as possible along with selecting the optimal intervention action a_t. Let μ be a policy which maps all past information $\mathcal{F}_t$ to the intervention action a_t. Denote by Π_{NA} the set of all non-anticipating policies. The quickest intervention problem is defined below:

PROBLEM 1 (QUICKEST INTERVENTION).

$$\min_{\mu \in \Pi_{NA}} C^\mu = \mathbb{E}_\mu \left[c_i^{a_0} + \sum_{t=1}^{T-1} \left(c_p^{z_t} + c_i^{a_t} \right) \right], \tag{9a}$$

$$s.t. \quad a_{t-1} \leq a_t \leq \min\{a_{t-1} + 1, A\}, \quad \forall 1 \leq t \leq T - 2. \tag{9b}$$

4 A POMDP APPROACH TO QUICKEST INTERVENTION

In this section, we view Problem 1 as a POMDP, and then convert it into a fully observed MDP by replacing the unobserved change-point indicator θ_t with the belief state π_t. With a grid approximation approach, we solve the optimal policy from the belief MDP.

4.1 Partially Observed MDP Formulation

The quickest intervention Problem 1 can be formulated as a partially observed MDP $\mathcal{M} = (\mathcal{S}, \mathcal{O}, \mathcal{A}, P(a), O(a), C(z, a))$ with the unknown indicator θ_t of the change-point τ. In particular, the POMDP is defined as follows:

- **State Space:** $S_t = (s_t, \theta_t, \tilde{a}_t) \in \{0, 1\}^2 \times \{0, 1, \cdots, A\}$. Here, s_t is the indicator whether the MDP has stopped; θ_t is the indicator whether the change point has happened; and $\tilde{a}_t$ is the current intervention level.
- **Observation Space:** $z_t \in \{1, \cdots, Z, S\}$. Here, z_t is the observation that the agent receives at time t, which helps the agent decide intervention actions. Notice that $z_t = S$ means the MDP has stopped.
- **Action Space:** $a_t \in \{0, 1, \cdots, A\}$. Action a_t is the intervention that the agent takes at time t to control the process.
- **State Transition Matrix:** At each time step, the change-point occurs with probability λ and the MDP stops with

probability $1 - \rho$, i.e.,

$$P_{s_t} = \begin{bmatrix} \rho & 1 - \rho \\ 0 & 1 \end{bmatrix}, P_{\theta_t} = \begin{bmatrix} 1 - \lambda & \lambda \\ 0 & 1 \end{bmatrix}, \tilde{a}_t = a_{t-1}.$$

- **Observation Model:** At each slot t, the agent observes whether the process has ended. If not, it receives an observation z_t following the distribution in Eq. (4).
- **One-step Cost:** With observation z_t and action a_t, the agent incurs a cost of $C(z_t, a_t) = c_p^{z_t} + c_i^{a_t}$.

It is well-known in [4] that solving a general POMDP is difficult. In the next subsection, we replace the unknown state dimension θ_t with its posterior π_t given all past information. Then, we convert the partially observed MDP into a fully observed MDP which is also called belief MDP.

4.2 Belief MDP Formulation

To convert a POMDP to an MDP, we replace θ_t with belief $\pi_t = \Pr(\theta_t = 1 | \mathcal{F}_t)$, where $\mathcal{F}_t$ is the σ-algebra that contains all past observations and actions, i.e., $\mathcal{F}_t = \{a_0, z_1, a_1, \cdots, a_{t-1}, z_t\}$. Suppose at time t, the posterior of $\theta_t = 1$ is π_t. Then, at the beginning of time slot $t + 1$, before receiving observation z_{t+1}, the posterior changes to $\tilde{\pi}_{t+1}$ as follows, due to the geometric prior distribution,

$$\begin{aligned} \tilde{\pi}_{t+1} &= \Pr(\theta_{t+1} = 1 | \mathcal{F}_t, a_t) \\ &\overset{(a)}{=} \Pr(\theta_t = 1 | \mathcal{F}_t)\Pr(\theta_{t+1} = 1 | \theta_t = 1, \mathcal{F}_t) \\ &\quad + \Pr(\theta_t = 0 | \mathcal{F}_t)\Pr(\theta_{t+1} = 1 | \theta_t = 0, \mathcal{F}_t) \\ &\overset{(b)}{=} \pi_t + \lambda(1 - \pi_t), \end{aligned} \tag{10}$$

where equality (a) holds because the change-point is independent of intervention a_t, and equality (b) is obtained from that the evolution of θ_t is a time-homogeneous Markov chain with transition matrix P_{θ_t}. Therefore, adding the observation z_{t+1} and action a_t into the σ-algebra, the update of the posterior π_{t+1} is as follows:

$$\begin{aligned} \pi_{t+1} &= \Pr(\theta_{t+1} = 1 | \mathcal{F}_{t+1}) \\ &\overset{(a)}{=} \frac{\Pr(\theta_{t+1} = 1 | \mathcal{F}_t, a_t)\Pr(z_{t+1} | \theta_{t+1} = 1, \mathcal{F}_t, a_t)}{\Pr(z_{t+1} | \mathcal{F}_t, a_t)} \\ &= \frac{\tilde{\pi}_{t+1} \beta_{a_t}(z_{t+1})}{\Pr(z_{t+1} | \mathcal{F}_t, a_t)}, \end{aligned} \tag{11}$$

where equality (a) follows from Bayes' rule. Note that from the total probability theorem, we have:

$$\begin{aligned} \Pr(z_{t+1} | \mathcal{F}_t, a_t) &= \Pr(\theta_{t+1} = 1 | \mathcal{F}_t, a_t)\Pr(z_{t+1} | \theta_{t+1} = 1, \mathcal{F}_t, a_t) \\ &\quad + \Pr(\theta_{t+1} = 0 | \mathcal{F}_t, a_t)\Pr(z_{t+1} | \theta_{t+1} = 0, \mathcal{F}_t, a_t) \\ &= \tilde{\pi}_{t+1} \beta_{a_t}(z_{t+1}) + (1 - \tilde{\pi}_{t+1})\alpha(z_{t+1}). \end{aligned} \tag{12}$$

Substituting Eq. (10) and Eq. (12) into Eq. (11), we conclude that the posterior can be updated iteratively as follows:

$$\begin{aligned} \pi_{t+1} &= T_{a_t}(\pi_t, z_{t+1}) \\ &= \frac{\tilde{\pi}_{t+1} \beta_{a_t}(z_{t+1})}{\tilde{\pi}_{t+1} \beta_{a_t}(z_{t+1}) + (1 - \tilde{\pi}_{t+1})\alpha(z_{t+1})}, \end{aligned} \tag{13}$$

where $\tilde{\pi}_{t+1} = \pi_t + \lambda(1 - \pi_t)$. The following theorem shows that the optimal policy for the quickest intervention problem can be obtained from studying the fully observed belief MDP.

THEOREM 4.1. *Solving the quickest intervention problem defined on the POMDP with $S_t = (s_t, \theta_t, \tilde{a}_t)$ and unknown state $\{\theta_t\}$ is equivalent to solving the problem on a fully observed MDP with $S_t = (\pi_t, \tilde{a}_t, z_t)$ where π_t replaces the unknown state θ_t. In the optimal solution, the intervention action a_t at slot t depends purely on the belief state π_t.*

The proof of Theorem 4.1 can be found in [19]. Based on it, the quickest intervention problem is converted to an MDP with the belief state π_t. To solve it, we define $V^t(\pi, \tilde{a}, z)$ to be the value function of state $S = (\pi, \tilde{a}, z)$ at time t, i.e.

$$V^t(\pi, \tilde{a}, z) = \min_{\mu \in \Pi_{NA}} \mathbb{E}_\mu \left[\sum_{k=t}^T C(z_k, \mu(\pi_k)) \,\middle|\, S_t = (\pi, \tilde{a}, z) \right]. \quad (14)$$

We further define $J_a^t(\pi, \tilde{a}, z)$ to be the action-value function of choosing $a_t = a$ at state $S = (\pi, \tilde{a}, z)$ at time t as follows.

$$J_a^t(\pi, \tilde{a}, z) = C(z_t, a)$$
$$+ \min_{\mu} \mathbb{E}_\mu \left[\sum_{k=t+1}^T C(z_k, \mu(\pi_k)) \,\middle|\, S_t = (\pi, \tilde{a}, z), a_t = a \right]. \quad (15)$$

For simplicity, in the following analysis, we set the value function $V^t(\pi, \tilde{a}, z)$ to be zero for any state that the MDP has stopped with stopping state $z = S$, i.e., $V^t(\pi, \tilde{a}, S) = 0$ for any π, $\tilde{a}$ and t. Also, we define $\sigma_a(\pi, z)$ to be the distribution of observation z_{t+1} given posterior π_t and action a_t as:

$$\sigma_a(\pi, z) = \Pr(z_{t+1} = z | \pi_t = \pi, a_t = a)$$
$$= \alpha(z)(1 - \tilde{\pi}) + \beta_a(z)\tilde{\pi}, \quad (16)$$

where $\tilde{\pi} = \pi + \lambda(1 - \pi)$ is the sudo-posterior before receiving the next observation z_{t+1}.

Recall that ρ is the probability the process continues. Therefore, the Bellman equation for action-value function $J_a^t(\pi, \tilde{a}, z)$ of the belief state MDP can be recursively expressed as follows:

$$J_a^t(\pi, \tilde{a}, z) = c_i^a + c_p^z + (1 - \rho)V^{t+1}(\pi, a, S)$$
$$+ \rho \sum_{z'=1}^Z \sigma_a(\pi, z')V^{t+1}(T_a(\pi, z'), a, z')$$
$$= c_i^a + c_p^z + \rho \sum_{z'=1}^Z \sigma_a(\pi, z')V^{t+1}(T_a(\pi, z'), a, z'). \quad (17)$$

Note that according to Eq. (9b), the agent can only choose a_t between $\tilde{a}_t$ and $\min\{\tilde{a}_t + 1, A\}$. The value function can also be expressed recursively as:

$$V^t(\pi, \tilde{a}, z) = \min \left\{ J_{\tilde{a}}^t(\pi, \tilde{a}, z), J_{\min\{\tilde{a}+1, A\}}^t(\pi, \tilde{a}, z) \right\}. \quad (18)$$

Since this MDP is a stochastic shortest path problem where the optimal policy is time-homogeneous and does not depend on time t [2], we omit the superscript t from now on. Notice that from Eq. (17), $J_a(\pi, \tilde{a}, z)$ is independent of $\tilde{a}$, and state z is not involved in deciding the optimal action. We can subtract the propagation cost c_p^z in the action-value function, and then remove state z and $\tilde{a}$ in $J_a^t(\pi, \tilde{a}, z)$. More specifically, denote $J_a(\pi) = J_a^t(\pi, \tilde{a}, z) - c_p^z$ to be a modified action-value function, and denote $V_{\tilde{a}}(\pi)$ to be the modified value function related to $J_a(\pi)$. Since they do not depend

on z, we will use z to replace z' in the remaining paper. Therefore, we can simplify the Bellman equation (17) and (18) as follows:

$$J_a(\pi) = c_i^a + \rho \sum_{z=1}^Z \sigma_a(\pi, z) \left[c_p^z + V_a(T_a(\pi, z)) \right], \quad (19a)$$

$$V_{\tilde{a}}(\pi) = \min\{J_{\tilde{a}}(\pi), J_{\min\{\tilde{a}+1, A\}}(\pi)\}. \quad (19b)$$

By solving the above Bellman equations, the agent obtains an optimal policy $\mu^*(\pi)$ for the quickest intervention problem defined in Problem 1. In the next subsection, we will prove the concavity and monotonicity of value functions. These properties will be useful in deriving a low-complexity algorithm in later sections.

4.3 A Structural Result for the Optimal Policy

We present an important property that the value functions of the belief MDP possess concavity and monotonicity.

LEMMA 4.2. *For any action a and current intervention level $\tilde{a}$, the value function $V_{\tilde{a}}(\pi)$ and the action-value function $J_a(\pi)$ are both concave and monotonically non-decreasing in belief state π. Furthermore, $V_A(\pi)$ and $J_A(\pi)$ are constant:*

$$V_A(\pi) = J_A(\pi) = \frac{c_i^A + \rho \sum_{z=1}^Z \alpha(z)c_p^z}{1 - \rho}. \quad (20)$$

The proof of Lemma 4.2 can be found in [19]. However, since π is continuous, directly solving the Bellman equations is generally difficult. In the next subsection, we present a grid-approximation approach to obtain a nearly optimal intervention policy through value iteration.

4.4 Solving Belief MDP through Grid Approximation

Algorithm 1 Grid Approximation Algorithm for $\mu^*(\pi)$

Require: grids $\{B_j\}_{1 \le j \le N}$, representatives $\{\pi_{B_j}\}_{1 \le j \le N}$, error tolerance ϵ.

1: **Initialization:** $k = 0$, $V_{\tilde{a}}^{(0)}(\pi_{B_j}) = 0$, $J_a^{(0)}(\pi_{B_j}) = 0$.
2: **while** $\exists B_j$ s.t. $|V_{\tilde{a}}^{(k)}(\pi_{B_j}) - V_{\tilde{a}}^{(k-1)}(\pi_{B_j})| > \epsilon$, **do**
3:     **for** $j = 1 : N$ **do**
4:         **for** $a = 0 : A$ **do**
5:             Update the action-value function $J_a^{(k+1)}(\pi_{B_j})$ according to $V_{\tilde{a}}^{(k)}(\pi_{B_j})$ and Eq. (19a).
6:         **end for**
7:     **end for**
8:     **for** $j = 1 : N$ **do**
9:         **for** $\tilde{a} = 0 : A$ **do**
10:             Update the value function $V_{\tilde{a}}^{(k+1)}(\pi_{B_j})$ and action $\mu(\pi_{B_j})$ according to $J_a^{(k+1)}(\pi_{B_j})$ and Eq. (19b).
11:         **end for**
12:     **end for**
13: **end while**
14: **for** $j = 1 : N$ **do**
15:     Set action $\mu(\pi)$ for any $\pi \in B_j$ to be $\mu(\pi_{B_j})$.
16: **end for**
Ensure: Nealy optimal policy $\{\mu(\pi)\}$.

In this section, we provide in Algorithm 1 a grid-approximation approach to solve the belief MDP [6, 9]. We first divide π into grids $\{B_1, B_2, \cdots, B_N\}$, where N is a proper number. One such design of grids may be uniform grids. Note that B_i's are non-intersected sections of real values satisfying $\cup_{j=1}^{N} B_j = [0, 1]$. Then, we select one element π_{B_i} in each grid B_j as its representative. The rest follows the standard value iteration [2].

Notice that this approach provides nearly optimal policies when the number of grids is large so that all grids are sufficiently small. However, since the performance of the obtained policy depends on the number of grids, the computational complexity is still very high when N is large. In the next section, we exploit linear approximation to obtain a low-complexity algorithm which has a nearly optimal provable guarantee in the *"local intervention"* regime.

5 A LOW-COMPLEXITY ALGORITHM WITH NEARLY PROVABLE GUARANTEE

In this section, we address the computational complexity of grid approximation by providing a heuristic low-complexity algorithm based on an approximated Bellman equation. We introduce the *"local intervention"* regime where the value function can be approximated accurately with the first-order Taylor's approximation. In this regime, we prove the nearly submodularity of the value function. Then, we design a threshold-based intervention policy, where the thresholds are easily calculable with closed-form expression.

5.1 "Local Intervention" Regime Analysis

To derive the low-complexity algorithm based on the Taylor's approximation, we consider the problem in the *"local intervention"* regime where all consecutive intervention levels are close to each other, and thus also close to the benign distribution. The formal definition of *"local intervention"* regime is presented in Assumption 1. The definition says that for a sufficiently small constant δ and any arbitrary discounted factor $\gamma \in (0, 1)$, the differences in cumulative distribution functions of $\boldsymbol{\beta}_a$'s for any state z are all very small.

ASSUMPTION 1. *("Local Intervention") For any action $a \geq 1$, the observation distribution $\beta_a(z)$ follows:*

$$\gamma\delta \leq \left| \sum_{z=j}^{Z} \beta_{a-1}(z) - \sum_{z=j}^{Z} \beta_a(z) \right| \leq \delta, \quad \forall j \geq 2, \tag{21}$$

where δ is a sufficiently small real value and $\gamma \in (0, 1)$.

This assumption implies that $\beta_a(z) - \alpha(z) = O(\delta)$ for any a. In later analysis, we will include superscript δ on any value that is related to δ. For the simplicity of notation, we define the following three values:

$$A_p^a = \sum_{z=1}^{Z} (\beta_a(z) - \alpha(z))c_p^z, \tag{22}$$

$$D_p^a = \sum_{z=1}^{Z} (\beta_a(z) - \beta_{a-1}(z))c_p^z, \tag{23}$$

$$D_i^a = c_i^a - c_i^{a-1}. \tag{24}$$

It is obvious to notice that $D_p^a = A_p^a - A_p^{a-1}$. According to the *"local intervention"* assumption, we have $A_p^a = \Theta(\delta)$ and $D_p^a = \Theta(\delta)$.

We then analyze the posterior update function Eq. (13) with Taylor's expansion evaluated at $\alpha(z)$ when δ is small enough:

$$
\begin{aligned}
T_a^\delta(\pi, z) &= \frac{\tilde{\pi}\beta_a(z)}{\tilde{\pi}\beta_a(z) + (1 - \tilde{\pi})\alpha(z)} \\
&= \frac{\tilde{\pi}(\alpha(z) + \beta_a(z) - \alpha(z))}{\tilde{\pi}(\alpha(z) + \beta_a(z) - \alpha(z)) + (1 - \tilde{\pi})\alpha(z)} \\
&= \tilde{\pi} + \frac{(1 - \tilde{\pi})\tilde{\pi}}{\alpha(z)}(\beta_a(z) - \alpha(z)) + O(\delta^2),
\end{aligned} \tag{25}
$$

where $\tilde{\pi} = \pi + \lambda(1 - \pi)$.

Substituting Eq. (25) into the expression of $J_a(\pi)$ in Eq. (19a), we obtain:

$$
\begin{aligned}
J_a^\delta(\pi) &= c_i^a + \rho \sum_{z=1}^{Z} [\alpha(z) + \tilde{\pi}(\beta_a(z) - \alpha(z))] \\
&\quad \cdot \left[c_p^z + V_a\left(\tilde{\pi} + \frac{(1 - \tilde{\pi})\tilde{\pi}}{\alpha(z)}(\beta_a(z) - \alpha(z)) + O(\delta^2) \right) \right] \\
&\stackrel{(a)}{=} \left[c_i^a + \rho \sum_{z=1}^{Z} \alpha(z)c_p^z \right] + \tilde{\pi}\rho A_p^a + \rho V_a(\tilde{\pi}) + O(\delta^2),
\end{aligned} \tag{26}
$$

where equality (a) is obtained by Taylor's expansion of $V_a(\cdot)$ evaluated at point $\tilde{\pi}$, as well as the definition (22). Also note that the first-order terms cancels due to $\sum_{z=1}^{Z} \alpha(z) = \sum_{z=1}^{Z} \beta_a(z) = 1$. Notice that the first term in bracket of Eq. (26) is independent of π, and the second term is linear, so in order to approximate the value function $J_a^\delta(\pi)$ linearly, we need to focus on the non-linear terms $\rho V_a(\tilde{\pi})$ and $O(\delta^2)$. Before we proceed, we first present the definitions of submodular functions and nearly-submodular functions.

Definition 5.1. An action-value function $J_a(\pi)$ is called submodular in pair (a, π) if it satisfies

$$\Delta_a(\pi) = J_{a+1}(\pi) - J_a(\pi)$$

is non-increasing with belief π.

From Topkis' lemma [16, Theorem 4.1], if a function $J_a(\pi)$ is submodular, the solution to its "argmin" problem, i.e.,

$$a^*(\pi) = \arg\min_a J_a(\pi),$$

is monotonically non-decreasing. Hence, a threshold exists since the action space is finite.

Next, we check the submodularity of our value function. For a given action $a < A$, subtract Eq. (26) for consecutive intervention actions and we get:

$$
\begin{aligned}
\Delta_a^\delta(\pi) &= J_{a+1}^\delta(\pi) - J_a^\delta(\pi) \\
&= \left[c_i^{a+1} - c_i^a \right] + \tilde{\pi}\left[\rho \sum_{z=1}^{Z}(\beta_{a+1}(z) - \beta_a(z))c_p^z \right] \\
&\quad + \rho\left[V_{a+1}^\delta(\tilde{\pi}) - V_a^\delta(\tilde{\pi}) \right] + O(\delta^2) \\
&= D_i^{a+1} + \tilde{\pi}\rho D_p^{a+1} + \rho\left[V_{a+1}^\delta(\tilde{\pi}) - V_a^\delta(\tilde{\pi}) \right] + O(\delta^2). \tag{27}
\end{aligned}
$$

Recall that in the *"local intervention"* regime, $\rho D_p^{a+1} = \Theta(\delta)$ is the same order as δ. Heuristically as δ is small enough, the higher order term $O(\delta^2)$ can be neglected. Therefore, we can accurately

approximate the difference $\Delta_a^\delta(\pi)$ in the *"local intervention"* regime through an approximated difference $\tilde{\Delta}_a^\delta(\pi)$ defined as follows:

$$\tilde{\Delta}_a^\delta(\pi) = D_i^{a+1} + \tilde{\pi}\rho D_p^{a+1} + \rho\left[\tilde{V}_{a+1}^\delta(\tilde{\pi}) - \tilde{V}_a^\delta(\tilde{\pi})\right], \quad (28)$$

where $\tilde{V}_a(\pi)$ and $\tilde{J}_a(\pi)$ are defined similarly by neglecting the higher orders in the original expression Eq. (26) as follows:

$$\tilde{J}_a(\pi) = \left[c_i^a + \rho \sum_{z=1}^Z \alpha(z)c_p^z\right] + \tilde{\pi}\rho A_p^a + \rho\tilde{V}_a(\tilde{\pi}), \quad (29a)$$

$$\tilde{V}_a(\pi) = \min\{\tilde{J}_a(\pi), \tilde{J}_{\min\{a+1,A\}}(\pi)\}. \quad (29b)$$

Eq. (29) can be interpreted as a mean approximation to the original Bellman equation in Eq. (19). Instead of computing the next-stage belief $T_a(\pi, z)$ and evaluating the corresponding value function $V_a(T_a(\pi, z))$ for each observation z as in Eq. (19a), we adopt the expected average belief update $\tilde{\pi}$ and substitute

$$\sum_{z=1}^Z \sigma_a(\pi, z)V_a(T_a(\pi, z))$$

with

$$V_a\left(\sum_{z=1}^Z \sigma_a(\pi, z)T_a(\pi, z)\right) = V_a(\tilde{\pi}).$$

With the approximated MDP defined in Eq. (29) and similar to submodularity, we introduce the definition of nearly-submodular functions in the *"local intervention"* regime.

Definition 5.2. We define an action-value function $J_a^\delta(\pi)$ to be nearly-submodular in pair (a, π) in the *"local intervention"* regime with δ, if its approximated difference function

$$\tilde{\Delta}_a^\delta(\pi) = \tilde{J}_{a+1}^\delta(\pi) - \tilde{J}_a^\delta(\pi)$$

defined in Eq. (28) is monotonically non-decreasing of π.

From the expression of $\tilde{\Delta}_a^\delta(\pi)$ above in Eq. (28), we prove the nearly-submodularity of $J_a^\delta(\pi)$ through induction of value iteration in the following Theorem 5.3.

THEOREM 5.3. *The belief MDP and its approximation with posterior π indicating whether change-point τ has occurred possess the following properties in the* "local intervention" *regime with δ:*

- *The action-value function $J_a^\delta(\pi)$ is nearly-submodular with pair (a, π).*
- *The policy $\tilde{\mu}^{*,\delta}(\pi)$ that solves the approximated Bellman equation (29) is monotonically non-decreasing, and thus possesses a threshold structure. That is, there exists a set of non-decreasing thresholds denoted as $\{\tilde{\pi}_a^{*,\delta}\}_{1\le a\le A}$, such that policy $\tilde{\mu}^{*,\delta}(\pi)$ switches to an action no less than a iff π exceeds threshold $\tilde{\pi}_a^{*,\delta}$.*

The proof of Theorem 5.3 can be found in [19]. From intuition, we expect that in the *"local intervention"* regime, the optimal policy $\mu^{*,\delta}(\pi)$ to the belief MDP behaves like a threshold-based policy just as $\tilde{\mu}^{*,\delta}(\pi)$, and the thresholds for the two policies will be very close to each other. This property will be identified in our numerical simulations.

Recall that the optimal cost C^{μ^*} of Problem 1 can be calculated from value function $V_0(0)$ defined in Eq. (19) with state $\tilde{a} = 0$ and

$\pi = 0$, i.e., $C^{\mu^*} = V_0(0)$. Similarly, we define an approximated optimal cost $\tilde{C}$ for the approximated MDP in Eq. (29) as follows:

$$\tilde{C} = \tilde{V}_0(0). \quad (30)$$

Intuitively, $\tilde{C}$ will also be close to C^{μ^*}. The following corollary provides a closed-form expression on $\tilde{C}$.

COROLLARY 5.4. *If all the thresholds $\{\tilde{\pi}_a^{*,\delta}\}_{1\le a\le A}$ are strictly within section $[0, 1)$, the total cost $\tilde{C}$ obtained by solving Bellman equation (29) is computed as follows:*

$$\tilde{C} = \sum_{a=1}^A \frac{\rho^{t_a}}{1-\rho}D_i^a + \sum_{a=1}^A \left(\frac{\rho^{t_a+1}}{1-\rho} - \frac{[\rho(1-\lambda)]^{t_a+1}}{1-\rho(1-\lambda)}\right)D_p^a$$
$$+ \frac{\rho\sum_{z=1}^Z \alpha(z)c_p^z}{1-\rho} + \frac{\lambda\rho A_p^0}{(1-\rho)[1-\rho(1-\lambda)]}. \quad (31)$$

where $t_a = \left\lceil \frac{\ln(1-\tilde{\pi}_a^{*,\delta})}{\ln(1-\lambda)} \right\rceil$ is the number of steps switching from action 0 to an action no less than a.

The proof of Corollary 5.4 can be found in [19]. However, calculating the thresholds $\{\tilde{\pi}_a^{*,\delta}\}_{1\le a\le A}$ is still computationally expensive. Therefore, we provide bounds on these thresholds in Corollary 5.5, which have close-form expressions.

COROLLARY 5.5. *The thresholds $\{\tilde{\pi}_a^{*,\delta}\}_{1\le a\le A}$ can be upper and lower bounded with closed-form expressions. Specifically, for a single threshold $\tilde{\pi}_a^{*,\delta}$ that $\tilde{\mu}^{*,\delta}(\pi)$ switches from $a-1$ to a, the following holds:*

$$\bar{\pi}_a + \frac{\sum_{j=a+1}^A \rho^{j-a}\left(D_i^j + \rho D_p^j\right)}{-(1-\lambda)\rho D_p^a} \le \tilde{\pi}_a^{*,\delta} \le \bar{\pi}_a, \quad (32)$$

where the upper bound threshold $\bar{\pi}_a$ is

$$\bar{\pi}_a = -\frac{D_i^a}{(1-\lambda)\rho D_p^a} - \frac{\lambda}{1-\lambda}. \quad (33)$$

The proof of Corollary 5.5 can be found in [19]. Heuristically, from Corollary 5.5, if the optimal policy $\mu^{*,\delta}(\pi)$ also possesses a threshold $\hat{\pi}_a^{*,\delta}$, it is highly likely that $\hat{\pi}_a^{*,\delta}$ lies in the same section or its $O(\delta)$ neighborhood when δ is small. On the other hand, the upper bound $\bar{\pi}_a$ serves as an approximated threshold which is easily calculable. In the next section, we present a low-complexity threshold-structured policy based on the upper bounds $\{\bar{\pi}_a\}_{1\le a\le A}$.

5.2 Low-complexity Policy based on Upper Bounds

Algorithm 2 Low-complexity Policy

1: Initialize the thresholds to be $\bar{\pi}_{A+1} \leftarrow 1$ and $\bar{\pi}_a \leftarrow \min\left\{-\frac{D_i^a}{(1-\lambda)\rho D_p^a} - \frac{\lambda}{1-\lambda}, \bar{\pi}_{a+1}\right\}$ for $a \le A$.
2: **while** MDP doesn't stop **do**
3:     Obtain observation z_t.
4:     Update belief $\pi_t = T(\pi_{t-1}, a_{t-1}, z_t)$.
5:     Implement action $a_t = a$, s.t. $\bar{\pi}_a \le \pi_t \le \bar{\pi}_{a+1}$.
6: **end while**

In this section, we propose our low-complexity policy $\bar{\mu}$. The policy and its implementation are summarized in Algorithm 2. The intuition is simple: since the upper bound $\bar{\pi}_a$ is very close to the original threshold $\tilde{\pi}_a^{*,\delta}$, choosing $\bar{\pi}_a$ to be the threshold is actually a conservative policy. In Sec 6, we show from the numerical simulations that the low-complexity policy's performance is nearly the same as the optimal one even beyond the *"local intervention"* regime.

6 SIMULATIONS

In this section, we first provide a lower bound on the total cost C^{μ^*} along with the definition of cost regret. These concepts are very important in our simulations. Then, we will proceed to present the numerical simulation results.

6.1 Lower Bound and Regret

Consider an oracle that knows the time horizon T and the change-point τ in advance. Then after relaxing the incremental constraint in Eq. (9b), the oracle can minimize the overall cost with a very simple oracle policy μ_o: choose $a_t = A$ when $\tau \leq t < T - 1$ and $a_t = 0$ otherwise. The "stricter is better" assumption in Eq. (8) ensures its optimality. We derive the cost of the oracle policy in the following Theorem 6.1.

THEOREM 6.1. *If the change point τ and horizon T is known ahead of time, then the optimal oracle action a_t is as follows:*

- *When $0 \leq t \leq \tau - 2$, $a_t = 0$.*
- *When $\tau - 1 \leq t \leq T - 2$, $a_t = A$.*
- *When $t = T - 1$, $a_t = 0$.*

The expected cost performance of oracle policy μ_o is

$$C^{\mu_o} = \frac{\rho \sum_{z=1}^{Z} \alpha(z) c_p^z}{1 - \rho} + c_i^A \left[\frac{\rho}{1 - \rho} - \frac{\rho(1 - \lambda)}{1 - \rho(1 - \lambda)} \right]. \quad (34)$$

The proof of Theorem 6.1 can be found in [19]. C^{μ_o} is the lower bound on the expected total cost, which only relates to system parameters such as ρ, λ, and $\boldsymbol{\alpha}$. To better understand the performance of our proposed policy, we study the cost regret for a given policy μ defined as follows:

Definition 6.2. The cost regret R^μ of a given policy μ is defined as the difference between its expected total cost C^μ and the expected cost C^{μ_o} of the oracle policy μ_o, i.e.,

$$R^\mu = C^\mu - C^{\mu_o}. \quad (35)$$

In the next subsection, we study the cost regret $R^{\bar{\mu}}$ of our proposed low-complexity policy and the cost regret R^{μ^*} of the optimal policy along with other benchmarks such as two-step detect-then-intervene policies modified from QCD [17].

6.2 Numerical Simulations

We compare the cost regret R^μ of the following four intervention policies, including our proposed low-complexity policy, optimal policy through grid approximation, and policies modified from quickest change detection algorithms [17]. These intervention policies are summarized as follows:

- **Low-complexity:** The low-complexity policy is proposed in Algorithm 2, which uses $\{\bar{\pi}_a\}$ as thresholds.

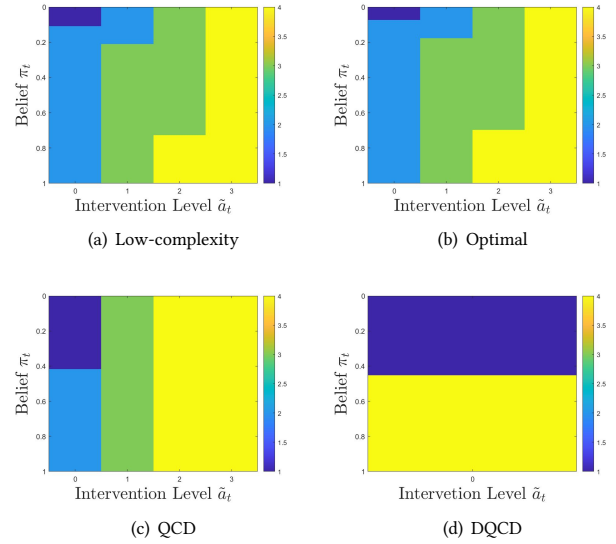(a) Low-complexity

(b) Optimal

(c) QCD

(d) DQCD

Figure 2: Illustration of intervention policies: (a) proposed low-complexity policy, (b) optimal policy, (c) quickest change detection policy, and (d) direct quickest change detection policy.

- **Optimal:** The optimal policy is by grid-approximation in Algorithm 1 with uniform and sufficiently small grids.
- **QCD:** The quickest change detection policy is a two-step intervention policy. It first detects the change point τ with Shiryaev's test [17] under an optimal chosen probability of false alarm (PFA). Then, it incrementally increases the intervention level to its maximum satisfying Eq. (9b).
- **Direct QCD:** The direct quickest change detection policy is also a two-step policy similar to QCD. The only difference is that it violates constraint Eq. (9b) and immediately implements the highest intervention level A once it claims the change-point has occurred.

Note that to verify the accuracy of approximations in the *"local intervention"* regime, we also compare the regret for approximated cost $\tilde{C}$ computed from Corollary 5.4.

We simulated the diffusion process with five observation states $z_t \in \{1, 2, 3, 4, 5\}$, where larger z_t represents a severer situation. The propagation costs associated with observations z_t are $[0, 1, 2, 3, 4]$, where $z_t = 1$ indicates a normal state where no propagation cost occurs. The actions which the agent can choose is $a_t \in \{0, 1, 2, 3\}$, where $a_t = 0$ represents idling and $a_t = 3$ indicates perfect intervention. The intervention costs are $[0, 0.02, 0.06, 0.2]$. As for the distribution of observations, we set the distribution before the change-point $\boldsymbol{\alpha}$ and that under the perfect intervention $\boldsymbol{\beta}_3$ to be $\boldsymbol{\alpha} = \boldsymbol{\beta}_3 = [0.2, 0.2, 0.2, 0.2, 0.2]'$. For distributions under other intervention actions, we set $\boldsymbol{\beta}_i = [0.2 - (6 - 2i)\delta, 0.2 - (3 - i)\delta, 0.3, 0.2 + (3 - i)\delta, 0.2 + (6 - 2i)\delta]$, $\forall i \in \{0, 1, 2, 3\}$.

We first studied the threshold structure of these policies and the results are summarized in Fig. 2. Recall that ρ is the geometric parameter that defines the distribution of horizon T, and λ is the
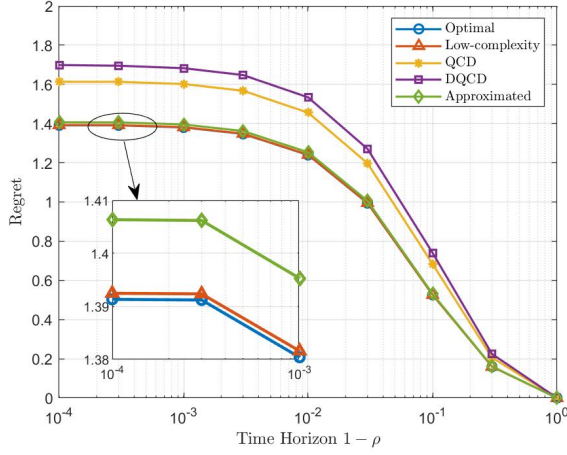
Figure 3: Cost regret performance as a function of time horizon parameter $1 - \rho$.

Figure 4: Cost regret performance as a function of change-point parameter λ. Note that DQCD performs even better than Optimal when $\lambda \to 1$ because it violates the constraint in Eq. (9b).

parameter that defines the geometric distribution of change-point τ. Here, we set $\rho = 0.99$ and $\lambda = 0.03$. We further chose $\delta = 0.02$. We see that the optimal policy possesses a threshold structure as we conjectured according to the results in the *"local intervention"* regime. Low-complexity also has a threshold very close to Optimal.

We then studied the performance of the four policies under different parameter ρ, the geometric parameter that defines horizon T. Larger $1 - \rho$ indicates the diffusion process stops more quickly. Here, we set $\lambda = 0.1$, $\delta = 0.02$ and then let $1 - \rho \in [10^{-4}, 1]$. The cost regret is depicted in Fig. 3. It shows that as ρ becomes smaller, the MDP stops with larger probability. Therefore, the probability that the agent needs to intervene also decreases. Low-complexity achieves a near-optimal regret performance under different ρ, and has much lower regret than the two QCD-based algorithms.

Next, we studied the performance of four policies under different λ, the parameter that defines the geometric distribution of change-point. Here, we set $\rho = 0.95$ and $\delta = 0.2$, and then let $\lambda \in [10^{-4}, 1]$. From Fig. 4, we can see that our proposed low-complexity policy achieves nearly optimal performance when λ is relatively large, but has a slightly larger regret than Optimal when λ is small. This is because under such a regime, the upper bound $\bar{\pi}_a$ becomes loose. With λ achieving 1, the change-point is more likely to happen at time 1. Since policy Direct QCD immediately switches to action A, violating the incremental increasing constraint in Eq. (9b), it achieves the best performance even better than Optimal.

Finally, we studied the cost performance of the four policies under different δ. Since δ is viewed as the distance between consecutive intervention action distributions, it can be replaced by the KL divergence of α and β_0 for better understanding. Here, we set $\rho = 0.95$, $\lambda = 0.1$, and let $\delta \in [10^{-4}, 0.03]$. The KL divergence $D(\alpha, \beta_0)$ lies within $[10^{-6}, 1]$. It shows that Low-complexity achieves nearly optimal performance under different δ. As a matter of fact, when the distributions are closer to each other, i.e., in the *"local intervention"* regime, the improvement of our low-complexity policy compared to QCD based policies becomes larger. The regret decreases by 22% when the KL divergence is as small as $O(10^{-6})$.
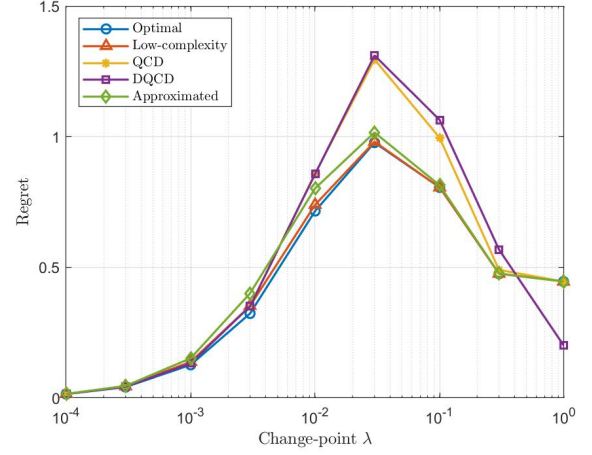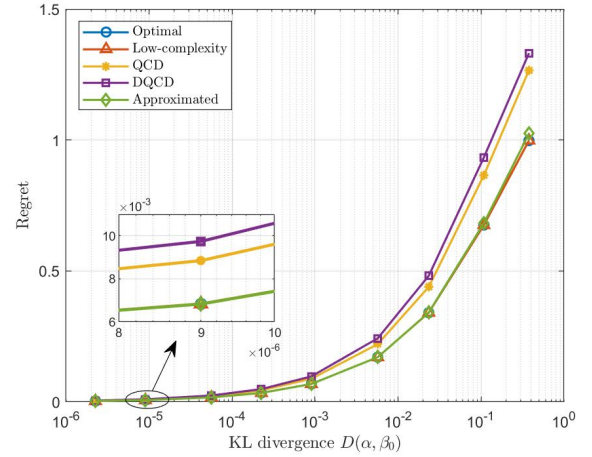

Figure 5: Cost regret performance as a function of KL divergence $D(\boldsymbol{\alpha}, \boldsymbol{\beta}_0)$.

Throughout the simulations, our approximated cost value $\tilde{C}$ is very close to the cost C^{μ^*} of the optimal policy. From Fig. 5, we see that the approximated cost regret only deviates from the regret of Optimal when the KL divergence $D(\alpha, \beta_0)$ deviates significantly from the *"local intervention"* regime, i.e., when $D(\alpha, \beta_0) \approx 0.5$. Otherwise, the approximation is very accurate.

6.3 Application to Anomaly Detection

In this subsection, we apply our low-complexity policy in algorithm 2 to a simple anomaly detection problem. Consider a scenario where a communication link is provided by a service provider to users for packets transmission. For simplicity, we assume at each
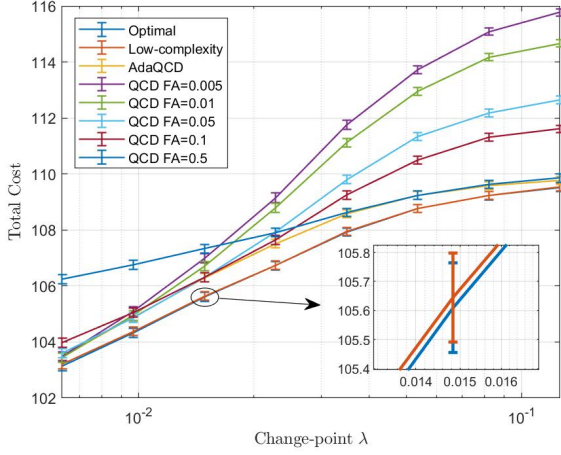
Figure 6: Total cost performance as a function of change-point parameter λ.

time t, a normal user chooses among five different service rates (denoted as $z_t \in \{1, 2, 3, 4, 5\}$ from low speed to high speed) to transmit these packets, equally likely (this assumption can be easily generalized). The user may become malicious after some change-points τ and start to send with an unusually high rate for a long period of time. The service provider monitors the user's transmission rate and charge the user accordingly. Based on the observations, the provider needs to decide whether the user has become malicious. If so, it may take intervention actions to throttle the data rate. The scenario described above can be viewed as a simplification of general MAC layer anomaly detection [3]. We assume that the service costs associated with different speeds, from low to high, are $\{0, 1, 2, 3, 4\}$ and the provider has four actions ($a_t \in \{0, 1, 2, 3\}$ where 0 represents idle) to choose from, which increasingly limit the high-speed transmission rate with intervention costs $\{0, 0.02, 0.06, 0.2\}$. The probability distributions of the service rate under different intervention actions are the same as in Sec. 6.2 with $\delta = 0.02$. However, we assume a fixed time horizon $T = 50$ in this simulation, and the service provider can choose any action at any time with no restriction such as Eq. (9b). In order to compute the low-complexity policy, we set $\rho = 0.98$. We simulated this system $20,000$ times and compared the performance of our low-complexity policy with the following baselines: (1) optimal policy computed through dynamic programming; (2) QCD policy with fixed PFA; (3) QCD policy with adaptively chosen optimal PFA. The total cost performance under different change-point parameter λ is shown in Fig. 6. It turns out that the total cost performance of our proposed low-complexity policy is nearly optimal, in the sense that its mean is close to the optimal policy and their confidence intervals are intersecting one another. A significant performance gap can be observed between the best QCD policy and the low-complexity policy, which demonstrates the superiority of our proposed policy. This result also shows that leveraging the assumptions such as Eq. (9b) in our technical sections will not harm the total cost performance of our proposed low-complexity policy too much, even though it is designed based on such assumptions.

7 CONCLUSION

We studied the problem of controlling a diffusion process with mutation. We formulated a POMDP and converted it into a fully observed MDP with belief states. We proposed a grid-approximation algorithm to compute the optimal intervention policy. To reduce its complexity, we considered the *"local intervention"* regime and approximated the value functions through first-order Taylor's expansion. We then proposed a low-complexity policy based on the upper bounds of belief thresholds derived from the approximation. Simulation results verify the performance of the proposed policies, which showed the low-complexity algorithm has a similar regret as the optimal policy even beyond the *"local intervention"* regime.

REFERENCES

[1] Taposh Banerjee, Miao Liu, and Jonathan P. How. 2017. Quickest change detection approach to optimal control in Markov decision processes with model changes. In *2017 American Control Conference (ACC)*. IEEE, USA, 399–405. https://doi.org/10.23919/ACC.2017.7962986

[2] Dimitri Bertsekas. 1995. *Dynamic Programming and Optimal Control*. Vol. 1. Athena, USA.

[3] Alvaro A. Cardenas, Svetlana Radosavac, and John S. Baras. 2009. Evaluation of Detection Algorithms for MAC Layer Misbehavior: Theory and Experiments. *IEEE/ACM Transactions on Networking* 17, 2 (2009), 605–617. https://doi.org/10.1109/TNET.2008.926510

[4] Vikram Krishnamurthy. 2016. *Partially Observed Markov Decision Processes*. Cambridge University Press, USA.

[5] Vikram Krishnamurthy and Udit Pareek. 2015. Myopic Bounds for Optimal Policy of POMDPs: An Extension of Lovejoy's Structural Results. *Oper. Res.* 63, 2 (April 2015), 428–434.

[6] Tze Siong Lau, Wee Peng Tay, and Venugopal V. Veeravalli. 2019. A Binning Approach to Quickest Change Detection With Unknown Post-Change Distribution. *IEEE Transactions on Signal Processing* 67, 3 (2019), 609–621. https://doi.org/10.1109/TSP.2018.2881666

[7] G. Lorden. 1971. Procedures for reacting to a change in distribution. *The Annals of Mathematical Statistics* 42, 6 (1971), 1897–1908. https://doi.org/10.1214/aoms/1177693055

[8] William S. Lovejoy. 1987. Some Monotonicity Results for Partially Observed Markov Decision Processes. *Oper. Res.* 35, 5 (Oct. 1987), 736–743.

[9] William S. Lovejoy. 1991. Computationally Feasible Bounds for Partially Observed Markov Decision Processes. *Operations Research* 39, 1 (1991), 162–175. https://doi.org/10.1287/opre.39.1.162

[10] George V. Moustakides. 1986. Optimal stopping times for detecting changes in distributions. *The Annals of Statistics* 14, 4 (1986), 1379–1387. https://doi.org/10.1214/aos/1176350164

[11] E. S. Page. 1954. Continuous Inspection Schemes. *Biometrika* 41, 1-2 (1954), 100–115. https://doi.org/10.1093/biomet/41.1-2.100

[12] Moshe Pollak. 1985. Optimal detection of a change in distribution. *The Annals of Statistics* 13, 1 (1985), 206–227.

[13] Y. Ritov. 1990. Decision theoretic optimality of the CUSUM procedure. *The Annals of Statistics* 18 (1990), 1464–1469.

[14] Albert N. Shiryaev. 1963. On optimum methods in quickest detection problems. *Theory of Probability & Its Applications* 8, 1 (1963), 22–46.

[15] Albert N. Shiryaev. 2007. *Optimal Stopping Rules*. Vol. 8. Springer Science & Business Media.

[16] Donald M. Topkis. 1978. Minimizing a Submodular Function on a Lattice. *Oper. Res.* 26, 2 (April 1978), 305–321. https://doi.org/10.1287/opre.26.2.305

[17] Venugopal V. Veeravalli and Taposh Banerjee. 2012. Quickest Change Detection. *arXiv* (2012).

[18] Honghao Wei, Xiaohan Kang, Weina Wang, and Lei Ying. 2019. QuickStop: A Markov Optimal Stopping Approach for Quickest Misinformation Detection. *Proceedings of the ACM on Measurement and Analysis of Computing Systems* 3, 2 (2019), 1–25.

[19] Qining Zhang, Honghao Wei, Weina Wang, and Lei Ying. 2022. On Low-Complexity Quickest Intervention of Mutated Diffusion Processes Through Local Approximation. *arXiv preprint arXiv:2206.04733* (2022).