Active Hypothesis Testing: Sequentiality and Adaptivity Gains

Mohammad Naghshvar and Tara Javidi
Department of Electrical and Computer Engineering,
University of California San Diego,
La Jolla, CA 92093 USA,
{naghshvar, tjavidi}@ucsd.edu

Abstract—Consider a decision maker who is responsible to collect observations so as to enhance his information in a speedy manner about an underlying phenomena of interest. The policies under which the decision maker selects sensing actions can be categorized based on the following two factors: i) sequential vs. non-sequential; ii) adaptive vs. non-adaptive. Non-sequential policies collect a fixed number of observation samples and make the final decision afterwards; while under sequential policies, the sample size is not known initially and is determined by the observation outcomes. Under adaptive policies, the decision maker relies on the previous collected samples to select the next sensing action; while under non-adaptive policies, the actions are selected independent of the past observation outcomes.

In this paper, performance bounds are provided for the policies in each category. Using these bounds, sequentiality gain and adaptivity gain, i.e., the gains of sequential and adaptive selection of actions are characterized.

Index Terms—Active hypothesis testing, performance bounds, feedback gain, reliability function.

I. INTRODUCTION

This paper considers a generalization of the classical hypothesis testing problem. Suppose there are \( M \) hypotheses among which only one is true. A Bayesian decision maker is responsible to enhance his information about the correct hypothesis in a speedy manner with a small number of samples while accounting for the penalty of wrong declaration.

In contrast to the classical \( M \)-ary hypothesis testing problem, at any given time, our decision maker can choose one of \( K \) available actions and hence, exert some control over the collected sample’s “information content.” We refer to this generalization as the active hypothesis testing problem. The special cases of active hypothesis testing naturally arise in a broad spectrum of applications in cognition [1], communications [2], anomaly detection [3], image inspection [4], group testing [5] and sensor management [6].

The sample size and the sensing actions can be selected either based on the past observation outcomes (on-line) or independent from them (off-line or open loop). According to this fact, the solutions are divided into four categories based on the following two factors: i) sequential vs. non-sequential; ii) adaptive vs. non-adaptive. Non-sequential schemes collect a fixed number of observation samples and make the final decision afterwards; while under sequential ones, the sample size is not set in advance and instead is determined by the specific observations made. Under adaptive policies, the decision maker relies on the previous collected samples to select the next sensing action; while under non-adaptive policies, the actions are selected independent of the past observation outcomes. A question of both theoretical and practical significance is the characterization of the benefits of making sequential and adaptive decisions relative to the non-sequential and non-adaptive solutions.

Due to the importance of the question, such gains have been characterized for many special cases of the active hypothesis testing [4], [7], [8], [9]. For instance, in [7] and [8], simple sequential and adaptive high dimensional reconstruction and sparse recovery are shown to significantly outperform the performance of the best non-sequential non-adaptive solutions. In contrast, [4] identifies scenarios where the gain in practice is insignificant. In this paper, we consider the problem of active hypothesis testing in its full generality and provide upper and lower bounds on the expected cost of the optimal sensing selection strategies in sequential and non-sequential as well as adaptive and non-adaptive classes of policies. The bounds are shown to be asymptotically tight (in terms of number of samples or equivalently in terms of reliability) and logarithmically increasing in the penalty of wrong declaration (or equivalently the error probability). As a simple corollary, we provide a full characterization of sequentiality and adaptivity gains in the general active hypothesis testing framework. These findings generalize and extend those of [7], and [8] by showing a logarithmic sequentiality gain in all cases and an additional logarithmic adaptivity gain in a large class of practically relevant cases. Furthermore, the results prove, as a corollary, the conjecture given in [4] on the insignificance of adaptivity gain when there exists a “most informative” sensing action which is independent of the Bayesian prior. Finally, we specialize our results in the active binary hypothesis testing case and state a simple necessary and sufficient condition for a logarithmic adaptivity gain. This parallels the results of [9] in the Bayesian context.

The remainder of this paper is organized as follows. In Section II, we formulate the problem and define various types of policies for selecting actions. Section III provides the main results of the paper and discusses the advantage of sequential
and adaptive selection of actions. Finally, in Section IV, active binary hypothesis testing is investigated as a special case and a necessary and sufficient condition for a logarithmic adaptivity gain is provided.

**Notations**: A random variable is denoted by an upper case letter (e.g. $X, Y, Z$) and its realization is denoted by a lower case letter (e.g. $x, y, z$). Similarly, a random column vector and its realization are denoted by bold face symbols (e.g. $X$ and $x$). For any set $S$, $|S|$ denotes the cardinality of $S$. For a set $A$, let $\Lambda(A)$ denote the collection of all probability distributions on elements of $A$, i.e., $\Lambda(A) = \{\lambda \in [0, 1]|A| : \sum_{a \in A} \lambda_a = 1\}$. The Kullback-Leibler (KL) divergence between two probability density functions $q(\cdot)$ and $q'(\cdot)$ is denoted by $D(q||q')$ where $D(q||q') = \int q(z) \log \frac{q(z)}{q'(z)} dz$. Finally, let $N(m, \sigma^2)$ denote a normal distribution with mean $m$ and variance $\sigma^2$.

**II. PROBLEM SETUP**

In Subsection II-A, we formulate the problem of active sequential hypothesis testing. Subsection II-B discusses different types of policies for selecting actions.

**A. Problem Formulation**

Here, we provide a precise formulation for the active $M$-ary sequential hypothesis testing problem.

Let $\Omega = \{1, 2, \ldots, M\}$. Let $H_i$, $i \in \Omega$, denote $M$ hypotheses of interest among which only one holds true. Let $\theta$ be the random variable that takes the value $\theta = i$ on the event that $H_i$ is true for $i \in \Omega$. We consider a Bayesian scenario where there is a uniform prior belief about $\theta$, i.e., $P(\theta = i) = 1/M$ for all $i \in \Omega$. $\mathcal{A}$ is the set of all sensing actions and is assumed to be finite with $|A| = K < \infty$. $\mathcal{Z}$ is the observation space. For all $a \in \mathcal{A}$, the observation kernel $q_i^a(\cdot)$ (on $\mathcal{Z}$) is the probability density function for observation $Z$ when action $a$ has been taken and $H_i$ is true. We assume that observation kernels $\{q_i^a(\cdot)\}_{a \in A}$ are known. Let $l$ denote the penalty for a wrong declaration, i.e., the penalty of selecting $H_j$, $j \neq i$, when $H_i$ is true. Let $\tau$ be the (stopping) time at which the decision maker retires. The objective is to find a sequence of sensing actions $A(0), A(1), \ldots, A(\tau - 1)$, a stopping time $\tau$, and a declaration rule $d : \mathcal{A}^\tau \times \mathcal{Z}^\tau \to \Omega$ that collectively minimize the expected total cost

$$
\mathbb{E}[\tau + 1_{\{d(A^\tau, Z^\tau) \neq \theta\}}],
$$

where the expectation is taken with respect to the initial belief as well as the distribution of observation sequence.

Note that in the above problem, the cost of a test is stated in terms of minimizing the expected sample size plus the expected penalty of wrong declaration. We are interested in the characterization of this cost as a function of penalty $l$. It is easy to show that under the optimal selection rule, the probability of error approaches zero as $l$ approaches infinity. Furthermore, as shown in [10], the above problem is (asymptotically) equivalent to the problem of minimizing the (expected) number of samples subject to a constraint $\epsilon = (l \log l)^{-1}$ on the expected probability of error.

**B. Types of Policies**

A policy is a rule based on which sensing actions $A(t)$, $t = 0, 1, \ldots, \tau - 1$ and stopping time $\tau$ are selected. The sensing actions and the stopping time can be selected either based on the past observation outcomes or independent from them. According to this fact, policies are divided into four categories based on the following two factors: i) sequential vs. non-sequential; ii) adaptive vs. non-adaptive. Non-sequential policies collect a fixed number of observation samples and make the final decision afterwards; while under sequential policies, the sample size is not known initially and is determined by the observation outcomes. More precisely, under non-sequential policies, $\tau = N$ for some $N \in \mathbb{N}$; while for sequential policies, $\tau$ is a random stopping time. Under adaptive policies, the decision maker relies on the previous collected samples to select the next sensing action; while under non-adaptive policies, the actions are selected independent of the past observation outcomes. We assume that under non-adaptive policies, sensing actions are selected according to a (randomized) rule $\lambda \in \Lambda(A)$ whose element $\lambda_a$ is the probability of selecting sensing action $a$.

**III. SEQUENTIALITY AND ADAPTIVITY GAINS**

In this section, we provide upper and lower bounds on the minimum expected total cost (1) under different types of policies defined in Subsection II-B. These bounds will be used then to characterize the gains of sequential and adaptive selection of actions.

We have the following technical Assumptions.

**Assumption 1.** For any two hypotheses $i$ and $j$, $i \neq j$, there exists an action $a$, $a \in \mathcal{A}$, such that $D(q_i^a||q_j^a) > 0$.

**Assumption 2.** $\max_{i \in \Omega} \max_{a \in \mathcal{A}} \max_{z \in \mathcal{Z}} q_i^a(z) < \infty$.

Assumption 1 ensures the possibility of discrimination between any two hypotheses. Assumption 2 implies that no two hypotheses are fully distinguishable using a single observation sample. Assumption 2 is for ease of our proofs. Standard techniques as in [11], [12] can be applied to generalize the bounds when Assumption 2 does not hold.

To continue with our analysis, we need the following definitions and notations.

**Definition.** For all $i \in \Omega$, $\lambda \in \Lambda(A)$, the reliability function of hypothesis $i$ with regard to randomized rule $\lambda$ is defined as

$$
R(i, \lambda) = \min_{j \neq i} \sum_{a \in \mathcal{A}} \lambda_a D(q_i^a||q_j^a),
$$

and the maximal randomized rule for hypothesis $i$ is denoted by

$$
\lambda^*_i = \arg \max_{\lambda \in \Lambda(A)} R(i, \lambda).
$$

This notion of reliability is a natural (and Bayesian) extension of reliability in classical detection where reliability function for hypothesis $i$ is related to type $i$ error probability.
For $\lambda \in \Lambda(A)$, let $\hat{R}(\lambda)$ denote the harmonic mean of $\{R(i, \lambda)\}_{i \in \Omega}$, i.e.,
\[
\hat{R}(\lambda) = \frac{M}{\sum_{i=1}^{M} \frac{1}{R(i, \lambda)}},
\]
and let $\hat{R}^*$ denote the harmonic mean of $\{R(i, \lambda^*_i)\}_{i \in \Omega}$, i.e.,
\[
\hat{R}^* = \frac{M}{\sum_{i=1}^{M} \frac{1}{R(i, \lambda^*_i)}}.
\]

Let $V_{NN}$, $V_{SN}$, and $V_{SA}$, denote the minimum expected total cost respectively under non-sequential non-adaptive, sequential non-adaptive, and sequential adaptive policies. Next we state the main results of the paper, i.e., performance bounds on the minimum expected total cost under the policies mentioned above.

**Proposition 1** (Non-sequential non-adaptive policy). Under Assumptions 1 and 2,\[
V_{NN} \geq \frac{2 \log l}{\max_{\lambda \in \Lambda(A)} \min_{i \in \Omega} R(i, \lambda)} - O(1). \tag{2}
\]

**Proposition 2** (Sequential non-adaptive policy). Under Assumptions 1 and 2, \[
V_{SN} = \frac{\log l}{\max_{\lambda \in \Lambda(A)} R(\lambda)} \pm O(1). \tag{3}
\]

**Proposition 3** (Sequential adaptive policy). Under Assumptions 1 and 2, \[
V_{SA} = \frac{\log l}{\hat{R}^*} \pm O(1). \tag{4}
\]

From Propositions 1-3, the minimum expected total cost under all considered policies grows logarithmically in $l$. However, the coefficient of the $\log l$ term is not the same in general and we have \[
\hat{R}^* \geq \max_{\lambda \in \Lambda(A)} \hat{R}(\lambda) \geq \max_{\lambda \in \Lambda(A)} \min_{i \in \Omega} R(i, \lambda). \tag{5}
\]

**Remark 1.** Generalizations of Propositions 1-3 are proved in the appendix. In particular, in the proof of Proposition 1, we provide an asymptotically tight lower and upper bound for $V_{NN}$. However, a simpler lower bound (2) is presented to avoid notational complexity.

Next we state and discuss the consequence of the bounds proposed above. In Subsection III-A, we focus on the advantage of causally selecting the retire/declaration time. In Subsection III-B, we discuss the gain of adaptively selecting sensing actions.

**A. Sequentiality Gain**

In this subsection, we discuss the advantage of causally selecting the retire/declaration time, i.e., $\tau$. In particular, we show that the performance gap between the sequential and non-sequential policy, $V_{NN} - V_{SN}$, grows logarithmically as the penalty $l$ increases. We refer to this performance gap as the sequentiality gain.

**Corollary 1.** The sequentiality gain is characterized as \[
V_{NN} - V_{SN} \geq \log l \left( \frac{2}{\max_{\lambda \in \Lambda(A)} \min_{i \in \Omega} R(i, \lambda)} - \frac{1}{\max_{\lambda \in \Lambda(A)} \hat{R}(\lambda)} \right) - O(1).
\]

**Remark 2.** The sequentiality gain grows logarithmically with $l$ and from (5), \[
V_{NN} - V_{SN} \geq \frac{\log l}{\max_{\lambda \in \Lambda(A)} \hat{R}(\lambda)} - O(1).
\]

**B. Adaptivity Gain**

In this subsection, the advantage of adaptively selecting the sensing actions is discussed. In particular, it is shown that the performance gap between the adaptive and non-adaptive policy, $V_{SN} - V_{SA}$, grows logarithmically as the penalty $l$ increases. We refer to this performance gap as the adaptivity gain.

**Corollary 2.** The adaptivity gain is characterized as \[
V_{SN} - V_{SA} = \log l \left( \frac{1}{\max_{\lambda \in \Lambda(A)} \hat{R}(\lambda)} - \frac{1}{\hat{R}^*} \right) \pm O(1).
\]

**Remark 3.** Unless there exists a $\hat{\lambda} \in \Lambda(A)$ such that, \[
R(i, \hat{\lambda}) = R(i, \lambda^*_i) \text{ for all } i \in \Omega,
\]
the adaptivity gain grows logarithmically with $l$.

Next we provide a sufficient condition under which there is no adaptivity gain. Before we proceed, we need the following definition and fact.

**Definition** (Blackwell Ordering [13]). Given two conditional probability densities $q^*$ and $q^b$ from $\mathcal{Z}$ to $\mathcal{Z}$, we say that $q^b$ is less informative than $q^*$ ($q^b \leq_B q^*$) if there exists a stochastic transformation $W$ from $\mathcal{Z}$ to $\mathcal{Z}$ such that\[
q^b_i(z) = \int q^*_i(y)W(y; z)dy \text{ for all } i \in \Omega. \tag{6}
\]

**Fact 1** (see [14]). If $q^b \leq_B q^*$, then for all $i, j \in \Omega$, \[
D(q^b_i || q^*_j) \leq D(q^b_i || q^*_i). \tag{7}
\]

**Corollary 3.** If there exists a sensing action $a^*$ satisfying $q^a \leq_B q^{a^*}$ for all $a \in A$, then there is no adaptivity gain.

**Remark 4.** The above corollary formalizes the notion of informativeness and confirms the conjecture provided in [4].

We close this section by a note on the class of non-sequential adaptive policies even though they seem rather unnatural to us (It is more reasonable to control the sample size using the observation outcomes if they are already being used to select sensing actions). Next proposition provides a lower bound on the minimum expected total cost under non-sequential adaptive policies, denoted by $V_{NA}$.

1Function $W : \mathcal{Z} \times \mathcal{Z} \to \mathbb{R}^+$ is called a stochastic transformation from $\mathcal{Z}$ to $\mathcal{Z}$ if it satisfies $\int_{\mathcal{Z}} W(y; z)dz = 1$. 

Proposition 4 (Non-sequential adaptive policy). Under Assumptions 1 and 2,
\[
V_{NA} \geq \min_{i \in \Omega} \max_{\lambda \in \Lambda(A)} \frac{\log l}{R(i, \lambda)} = O(1). \quad (7)
\]

IV. SPECIAL CASE: BINARY HYPOTHESIS TESTING

In this section, we consider active binary hypothesis testing \((M = 2)\) as a special case. The performance bounds provided in Section III are simplified by substituting the following equations into the denominators of the bounds.

\[
R(1, \lambda) = \sum_{a \in A} \lambda_a D(q_1^a || q_2^a), \quad R(2, \lambda) = \sum_{a \in A} \lambda_a D(q_2^a || q_1^a),
\]
\[
R(1, \lambda_1^*) = \max_{a \in A} \sum_{a \in A} \lambda_a D(q_1^a || q_2^a), \quad R(2, \lambda_2^*) = \max_{a \in A} \sum_{a \in A} \lambda_a D(q_2^a || q_1^a),
\]
\[
\bar{R}(\lambda) = \left( \sum_{a \in A} 0.5 \lambda_a D(q_1^a || q_2^a) + \sum_{a \in A} 0.5 \lambda_a D(q_2^a || q_1^a) \right)^{-1},
\]
\[
\bar{R}^* = \left( \max_a 0.5 \sum_{a \in A} \lambda_a D(q_1^a || q_2^a) + \max_a 0.5 \sum_{a \in A} \lambda_a D(q_2^a || q_1^a) \right)^{-1}.
\]

Next we state a simple necessary and sufficient condition for a logarithmic adaptivity gain in the active binary hypothesis testing case.

Corollary 4. In the active binary hypothesis testing case, the adaptivity gain grows logarithmically in \(t\) if and only if
\[
\arg \max_{a \in A} D(q_1^a || q_2^a) \neq \arg \max_{a \in A} D(q_2^a || q_1^a).
\]

A. Numerical Example

Consider the active binary hypothesis testing problem with additive Gaussian noisy observations under two actions \(a\) and \(b\) shown in Fig. 1. In this example, the observation noise associated with actions \(a\) and \(b\) are such that they add unequal noise to the hypotheses. In the remainder of this subsection, we compare the performance of all considered policies for this example.

![Diagram of sensing actions](image)

Fig. 1. Active binary hypothesis testing problem with additive Gaussian noisy observations.

Table I compares the performance bounds of the considered policies for the example of Fig. 1.

<table>
<thead>
<tr>
<th>Table I</th>
<th>COMPARISON OF PERFORMANCE BOUNDS FOR THE EXAMPLE OF FIG. 1.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sequential</td>
</tr>
<tr>
<td>Adaptive</td>
<td>log (l/2.98)</td>
</tr>
<tr>
<td>Non-adaptive</td>
<td>log (l/2.27)</td>
</tr>
</tbody>
</table>

APPENDIX

In this appendix, generalizations of Propositions 1-4 are proved. Before we proceed, we need the following definition, lemma, and facts.

Definition. The Rényi divergence of order \(\alpha\) between two probability density functions \(q(\cdot)\) and \(q'(\cdot)\) is denoted by \(D_\alpha(q||q')\) where \(D_\alpha(q||q') = \frac{1}{\alpha-1} \log \int q^{\alpha}(x)q'^{-\alpha}(x)dx\).

Lemma 1. For two probability density functions \(q(\cdot)\) and \(q'(\cdot)\) and for all \(\alpha \in [0, 1]\) we have
\[
(1 - \alpha)D_\alpha(q||q') \leq \min \left\{ \{1 - \alpha\}D(q||q'), \alpha D(q'||q) \right\}.
\]

Fact 2 (Kolmogorov’s Maximal Inequality [15]). Suppose \(X_t\) for \(t = 1, 2, \ldots, M\) be independent random variables with \(\mathbb{E}[X_t] = 0\) and \(\text{Var}(X_t) < \infty\). Let \(S_n = \sum_{t=1}^{n} X_t\). Then
\[
P \left( \max_{0 \leq n \leq N} |S_n| > x \right) \leq \frac{\text{Var}(S_N)}{x^2} = \sum_{t=1}^{N} \frac{\text{Var}(X_t)}{x^2}.
\]

Fact 3 (McDiarmid’s Inequality [16]). Let \(X = (X_1, \ldots, X_n)\) be a family of independent random variables with \(X_k\) taking values in a set \(\mathcal{X}_k\) for each \(k\). Suppose a real-valued function \(f\) defined on \(\Pi_k \mathcal{X}_k\) satisfies
\[
|f(x) - f(x')| \leq c_k,
\]
whenever the vectors \(x\) and \(x'\) only differ in the \(k\)-th coordinate. Then for any \(\nu > 0\),
\[
P(f(X) - \mathbb{E}[f(X)] \geq \nu) \leq e^{-2\nu^2 \sum_{k=1}^{n} c_k^2},
\]
\[
P(f(X) - \mathbb{E}[f(X)] \leq -\nu) \leq e^{-2\nu^2 \sum_{k=1}^{n} c_k^2}.
\]

A. Information State as Sufficient Statistic

The problem of active \(M\)-ary hypothesis testing is a partially observable Markov decision problem (POMDP) where the state is static and observations are noisy. It is known that any POMDP is equivalent to an MDP with a compact yet uncountable state space, for which the belief of the decision maker about the underlying state becomes an information state [17]. In our setup, thus, the information state at time \(t\) is nothing but a belief vector specified by the conditional probability of hypotheses \(H_1, H_2, \ldots, H_M\) to be true given the initial belief and all the previous observations and actions. Accordingly, the information state space is defined as \(\mathcal{P}(\Theta) = \{\rho \in [0, 1]^M : \sum_{i=1}^{M} \rho_i = 1\}\) where \(\Theta\) is the \(\sigma\)-algebra generated by random variable \(\theta\). In one sensing step, the evolution of the belief vector follows Bayes’ rule.
Let $\rho(n)$ denote the posterior belief after $n$ observations. The expected total cost (1) can be rewritten as
\[ \mathbb{E}[\tau] + \ell \bar{P} \epsilon, \]  
where $\bar{P} \epsilon = \mathbb{E}[1 - \max_{j \in \Omega} \rho_j(\tau)]$ is the probability of wrong declaration.

Let $V_{NN}(\rho)$, $V_{SN}(\rho)$, $V_{SA}(\rho)$, $V_{NA}(\rho)$, denote the minimum expected total cost respectively under non-sequential non-adaptive, sequential non-adaptive, sequential adaptive, and non-sequential adaptive policies for prior belief $\rho$. Next we prove Propositions 1-4 for general prior belief $\rho$.

B. Proposition 1, non-sequential non-adaptive policy

In this subsection, we show that
\[ V_{NN}(\rho) \leq \frac{\log l + \log(M - 1) - \min_{i,j \in \Omega} \frac{\rho_i}{\rho_j}}{\hat{D}} + O(1), \]  
(9)
\[ V_{NN}(\rho) \geq \frac{\log l - \max_{i,j \in \Omega} \frac{\rho_i}{\rho_j}}{\hat{D}} - O(1), \]  
(10)
where
\[ \hat{D} = \max_{\lambda \in \Lambda(\mathcal{A})} \min_{i \neq j} \min_{a \in [0,1]} \sum_{a \in \mathcal{A}} \lambda_a (1 - \alpha) D_a(q^a_i || q^a_j). \]  
(11)

From Lemma 1,
\[ \hat{D} \leq 0.5 \max_{\lambda \in \Lambda(\mathcal{A})} \min_{i \neq j} \min_{a \in \mathcal{A}} \sum_{a \in \mathcal{A}} \lambda_a D(q^a_i || q^a_j). \]  
(12)
Combining (10),(12), we have the assertion of Proposition 1.

Proof:

Suppose $\hat{\lambda} \in \Lambda(\mathcal{A})$ achieves the maximum in (11). Let $\pi_{NN}$ be a non-sequential non-adaptive policy that collects
\[ \hat{n} = \left( \log l + \log(M - 1) - \min_{i,j \in \Omega} \frac{\rho_i}{\rho_j} \right) / \hat{D} \]  
(13)
observation samples and selects sensing actions according to the randomized rule $\hat{\lambda}$. The expected total cost under this policy is $\hat{n} + \ell \bar{P} \epsilon$. Next we find an upper bound for $\bar{P} \epsilon$. Let $Z_i = \{ Z^a : \rho_i(\hat{n}) \geq \rho_j(\hat{n}) \}$ for all $i \in \Omega$ and $e_{ij} = P(\{ Z^a : \rho_i(\hat{n}) < \rho_j(\hat{n}) \} | \theta = i)$.
\[ \bar{P} \epsilon = \sum_{i=1}^{M} \rho_i P(Z_i | \theta = i) \leq \sum_{i=1}^{M} \rho_i \sum_{j \neq i} e_{ij} \leq (M - 1) \max_{i,j \in \Omega} e_{ij}. \]  
(14)
Following similar lines as the proof of Theorem 9 in [18], we can show that for all $i, j \in \Omega$, and $\alpha \in [0,1]$,
\[ \max \{ e_{ij}, e_{ji} \} \leq \exp \left( -\hat{n}(1 - \alpha) \sum_{a \in \mathcal{A}} \lambda_a D_a(q^a_i || q^a_j) - \min_{i,j} \frac{\rho_i}{\rho_j} \right). \]  
(15)
Combining (11),(13)-(15), we obtain $\bar{P} \epsilon = O(\frac{1}{D})$ and hence,
\[ V_{NN} \leq \hat{n} + \ell \bar{P} \epsilon = \hat{n} + O(1). \]

This completes the proof of upper bound. Next the proof of lower bound is given.

Consider a policy $\pi_{SN}$ that collects $n$ observation samples according to $\lambda \in \Lambda(\mathcal{A})$. From (14) we have
\[ \bar{P} \epsilon \geq \rho_i e_{ij} + \rho_j e_{ji} \quad \text{for any } i, j \in \Omega. \]  
(16)
Moreover, it can be shown that for all $i, j \in \Omega$, and $\alpha \in [0,1]$,
\[ \max \{ e_{ij}, e_{ji} \} \geq \exp \left( -n(1 - \alpha) \sum_{a \in \mathcal{A}} \lambda_a D_a(q^a_i || q^a_j) - \max_{i,j} \frac{\rho_i}{\rho_j} \right). \]  
(17)
Combining (16),(17), a lower bound is obtained for the expected total cost under policy $\pi_{SN}$. The lower bound for $V_{SN}$ is obtained by minimizing over the choice of $n$ and $\lambda$.

C. Proposition 2, sequential non-adaptive policy

In this subsection, we show that
\[ V_{SN}(\rho) \leq \min_{\lambda \in \Lambda(\mathcal{A})} \sum_{i=1}^{M} \log l + \log M - \min_{k \neq i} \frac{\rho_k}{\rho_i} R(i,\lambda) + O(1), \]  
(18)
\[ V_{SN}(\rho) \geq \min_{\lambda \in \Lambda(\mathcal{A})} \sum_{i=1}^{M} \log l - \max_{k \neq i} \frac{\rho_k}{\rho_i} R(i,\lambda) - O(1). \]  
(19)

Proof:

Suppose $\hat{\lambda} \in \Lambda(\mathcal{A})$ achieves the minimum in (18). The upper bound is achieved by a policy that selects sensing actions according to $\hat{\lambda}$ and stops sampling at $\tau = \min\{ n : \max_{i \in \Omega} \rho_i(n) \geq 1 - l^{-1} \}$. The proof follows similar lines as the proof of Proposition 2 in [19]. Next the proof of lower bound is provided.

Consider arbitrary $0 < \epsilon < 1$ and $\delta > 0$. Let $\pi_{SN}$ denote a sequential policy that selects sensing actions according to $\lambda \in \Lambda(\mathcal{A})$ and stops sampling whenever $\bar{P} \epsilon \leq \epsilon$. For all $i \in \Omega$, let $\tau_i = \min\{ n : \rho_i(n) \geq 1 - \epsilon \}$ and
\[ T_i = \log \frac{1 - \epsilon}{1 - \epsilon - \max_{k \neq i} \frac{\rho_k}{\rho_i} R(i,\lambda + \delta)}. \]  
(20)

The expected total cost under policy $\pi_{SN}$ is lower bounded as
\[ \mathbb{E}[\tau] + \ell \bar{P} \epsilon \geq \sum_{i=1}^{M} \rho_i(T_i + \epsilon T_i) P(\tau_i > T_i | \theta = i). \]  
(21)

By Assumption 2, there exists $c, c < \infty$, such that
\[ \max_{i,j \in \Omega} \sup_{a \in \mathcal{A}} \frac{\rho_i(a)}{\rho_j(a)} \leq c. \]

Using Fact 2, we can show that
\[ P(\tau_i < T_i | \theta = i) \leq \frac{T_i c^2}{(T_i \delta)^2} = \frac{c^2}{T_i \delta^2}. \]  
(22)
Combining (20)-(22) a lower bound is obtained for the expected total cost under policy $\pi_{SN}$. The lower bound for $V_{SN}$ is obtained by minimizing over the choice of $\epsilon$ and $\lambda$.

\[ \]
D. Proposition 3, sequential adaptive policy

We have
\[
V_{SA}(\rho) \leq \sum_{i=1}^{M} \rho_i \frac{\log l + \log(M - 1) - \min_{k \neq i} \log \frac{\rho_k}{\rho_i}}{R(i, \lambda_i^*)} + O(1),
\]
\[
V_{SA}(\rho) \geq \sum_{i=1}^{M} \rho_i \frac{\log l - \max_{k \neq i} \log \frac{\rho_k}{\rho_i}}{R(i, \lambda_i^*)} - O(1).
\]

The upper bound was proved in [19]. The proof of lower bound simply follows similar lines as the proof of Proposition 2 where \(R(i, \lambda)\) is replaced by \(R(i, \lambda_i^*)\).

E. Proposition 4, non-sequential adaptive policy

In this subsection, we show that
\[
V_{NA}(\rho) \geq \frac{\log l - \max_{k \neq i} \log \frac{\rho_k}{\rho_i}}{\min_{i \in \Omega} \max_{\lambda \in \Lambda(A)} R(i, \lambda)} - O(1).
\]

Proof:

Let \(\pi_{NA}\) be a non-sequential adaptive policy that collects \(n\) observation samples. Consider an arbitrary \(\delta > 0\) and let
\[
\epsilon_i = \frac{1}{\exp \left( \left\{ n(R(i, \lambda_i^*) + \delta) + \max_{k \neq i} \log \frac{\rho_k}{\rho_i} \right\} \right) + 1}.
\]

We have
\[
\tilde{P} \geq \sum_{i=1}^{M} \rho_i \mathbb{E}[1 - \rho_i(n)\theta = i] P(Z_i|\theta = i),
\]
where
\[
\mathbb{E}[1 - \rho_i(n)\theta = i] \geq \epsilon_i P(1 - \rho_i(n) \geq \epsilon_i | \theta = i).
\]

Let \(\hat{j} = \arg\min_{j \neq i} \sum_{t=0}^{n-1} \mathbb{E}[\log \frac{q_j(t)}{q_i(t)}|Z_j|\theta = i]\) where actions \(\{A(t)\}_{t=0}^{n-1}\) are selected according to \(\pi_{NA}\).

\[
P(1 - \rho_i(n) < \epsilon_i | \theta = i) \leq P \left( \frac{\rho_i(n)}{1 - \rho_i(n)} > \log \frac{1 - \epsilon_i}{\epsilon_i} | \theta = i \right)
\]
\[
\leq P \left( \log \frac{\rho_i(n)}{\rho_j(n)} > \log \frac{1 - \epsilon_i}{\epsilon_i} \right) | \theta = i
\]
\[
\leq P \left( \log \frac{\rho_i(n)}{\rho_j(n)} - \mathbb{E}[\log \frac{\rho_i(n)}{\rho_j(n)}] > \log \frac{1 - \epsilon_i}{\epsilon_i} - \max_{k \neq i} \log \frac{\rho_k}{\rho_i} - nR(i, \lambda_i^*) \right) | \theta = i
\]
\[
\leq P \left( \log \frac{\rho_i(n)}{\rho_j(n)} - \mathbb{E}[\log \frac{\rho_i(n)}{\rho_j(n)}] > n\delta | \theta = i \right)
\]
\[
\leq \exp(-n\delta^2/c^2),
\]
where (a) follows from the fact that given \(\{\theta = i\}\),
\[
\mathbb{E}[\log \frac{\rho_i(n)}{\rho_j(n)}] = \log \frac{\rho_i(t + 1)}{\rho_j(t + 1)} + \sum_{t=0}^{n-1} \mathbb{E}[\log \frac{q^g_{A(t)}(Z)| \theta = i}{q^g_{j(t)}(Z)}]
\]
\[
= \log \frac{\rho_i(t + 1)}{\rho_j(t + 1)} + \sum_{t=0}^{n-1} \mathbb{E}[\log \frac{q^g_{A(t)}(Z)| \theta = i}{q^g_{j(t)}(Z)}]
\]
\[
\leq \max_{k \neq i} \log \frac{\rho_k}{\rho_i} + \min_{j \neq i} \lambda^*_a(\lambda | \theta = i) D(\pi_{NA}^a||q^g_j),
\]
and (b) follows from Fact 3.

Similarly, it can be shown that
\[
P(Z_i^c|\theta = i) \leq \exp(-n(R(i, \lambda_i^*))^2/c^2).
\]

Combining (23)-(26) and minimizing the bound over \(n\), we have the assertion of the proposition.

REFERENCES