2016 IEEE 55th Conference on Decision and Control (CDC)
ARIA Resort & Casino
December 12-14, 2016, Las Vegas, USA

Information Acquisition and Sequential Belief Refinement

Tara Javidi
Electrical & Computer Engineering
University of California, San Diego
Email: tara@ece.ucsd.edu

***Abstract*—Information acquisition and utilization problems (IAUP) form a class of stochastic decision problems in which a decision maker is faced with utilizing a stochastically varying (and uncontrollable) environment. However, the state of the environment, due to the limited nature of the measurements in terms of dimension/cost/accuracy, is only partially known to the decision maker. The decision maker, by carefully controlling the sequence of actions with uncertain outcomes and noisy measurements, dynamically refines the belief about the stochastically varying parameters of interest. A generalization of hidden Markov models and a special case of partially observable Markov models, information acquisition is both an informational problem as well as a control one.**

In this tutorial, we start with the stochastic control view of the problem and show that the class of information acquisition and utilization problems is equivalent to generalized dynamic tracking. We also state and discuss the appropriate dynamic programming formulation and the insight it provides. In the next section of the tutorial, we focus on active hypothesis testing as an important special case of information acquisition problems. This problem has been studied in various areas of applied mathematics, statistics, and engineering. After a short discussion on the historical developments due to Wald, Blackwell, DeGroots, and Chernoff, we catalog recent advances to connect DeGroot's information utility framework with the Shannon theoretic concept of uncertainty reduction to introduce a symmetrized divergence measure: Extrinsic Jensen-Shannon (EJS) divergence.

In the last part of tutorial, we visit two special cases of the problem: Noisy Search and Bayesian Active Learning. In both, we use the EJS divergence to propose a new learning/search strategy and develop performance guarantees. Obtaining (tight) lower and upper bounds on the optimal performance, as a corollary, proves the significant (asymptotic) performance gain of sequential and adaptive search strategies over open-loop ones.

***Index Terms*—Information Acquisition and Utilization, Partially Observable MDP, Active Hypothesis Testing.**

I. INTRODUCTION

This paper focuses on the problem of information acquisition and utilization where a decision maker, by carefully controlling a sequence of actions with uncertain outcomes, dynamically refines his/her belief about stochastically time-varying parameters of interest in order to utilize the system as efficiently as possible.

This work was partially supported by National Science Foundation Grants: CCF-1513883, CNS-1329819, AST-1247995, and CCF-1302588. The author would like to acknowledge her current PhD students S. Chiu and A. Lalitha for their generous help as well as her former PhD student, Dr. M. Naghshvar, whose dissertation laid out much of what is discussed in this tutorial paper.

This tutorial discusses a new theoretical framework for stochastic learning and decision-making in such a setting termed *Information Acquisition and Utilization Problems* (IAUP). IAUP is a special case of partially observable Markov decision problems (POMDP) [1] with several unique properties. First, in an IAUP, there is a twin-set of actions explicitly corresponding to the joint activities of acquisition (typically corresponding to sensing, sampling, measurement, etc) and utilization (typically corresponding to transmission, cost minimization, etc). While, this explicitly allows for different costs to be associated with different acquisition choices, reflecting how much energy/bandwidth is needed to obtain a particular granularity and accuracy of information, the twin-actions in our IAUP formulation do not affect the underlying stochastic evolution of the environment (the time-varying stochastic process). In other words, while the problem clearly fits in the general framework of stochastic control, it is an entirely informational problem with a real-time flavor (tracking states in a dynamic fashion). In particular, the first contribution of this work is to establish the general attributes of IAUPs in form of its average cost optimality equation (ACOE) and its equivalence to a generalized real-time tracking problem. This result concretely confirms our original intuition that IAUPs are, unlike most general stochastic control problems POMDP, purely *informational* problems.

The second part of the tutorial focuses on a synthesis of the prior works on active hypothesis testing [2]. Active hypothesis testing has been studied in various areas of applied mathematics, statistics, and engineering. After a short discussion on the historical developments due to Wald, Blackwell, DeGroot's, and Chernoff, we catalog recent advances to connect DeGroot's information utility framework with the Shannon theoretic concept of uncertainty reduction to introduce a symmetrized divergence measure: Extrinsic Jensen-Shannon (EJS) divergence. We then investigate the performance of a heuristic based this divergence in the context of the two special cases of noisy dynamic search [3], and Bayesian Active Learning with non-Persistent Noise [4].

Notation: Let $[x]^+ = \max\{x, 0\}$. We use boldface letters to represent vectors. And we write $\boldsymbol{\rho}^{\downarrow}$ to denote sorted element of a vector $\boldsymbol{\rho}$ in descending order, $i.e.$, $\rho_i^{\downarrow}$ represents the ith largest element of $\boldsymbol{\rho}$. For any set $\mathcal{S}$, $|\mathcal{S}|$ denotes the cardinality of $\mathcal{S}$. The space of all probability distributions on set $\mathcal{A}$ is denoted by $\mathbb{P}(\mathcal{A})$. The filtration

$\{\mathcal{F}_t\}$ denotes the nested σ-algebra associated with random sequence $Y_{1:t-1}$. All logarithms are in base 2. The entropy function on a vector $\boldsymbol{\rho} = [\rho_1, \rho_2, \ldots, \rho_M] \in [0,1]^M$ is defined as $H(\boldsymbol{\rho}) = \sum_{i=1}^{M} \rho_i \log \frac{1}{\rho_i}$, the Kullback-Leibler (KL) divergence between distribution P and Q is denoted by $D(P\|Q) = \sum_x P(x) \log \frac{P(x)}{Q(x)}$ and the mutual information between random variable X and Y is defined as $I(X,Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$, where $p(x,y)$ is the joint distribution and $p(x)$ and $p(y)$ are the marginals of X and Y with the convention $0 \log \frac{a}{0} = 0$ and $b \log \frac{b}{0} = \infty$ for $a, b \in [0,1]$ with $b \neq 0$. Let Bern(p) denote the Bernoulli distribution with parameter p, and $I(q, p[q])$ denote the mutual information of the input $X \sim$ Bern(q) and the output Y of a BSC channel with crossover probability $p[\alpha]$. $C(\alpha) := I(1/2, p(\alpha))$ denotes the capacity of a BSC with crossover probability $p[\alpha]$ and $C_1(\delta) := D(\text{Bern}(p[\delta]) \,\|\text{Bern}(1-p[\delta]))$ denotes the best possible error exponent of adaptive sequential search with search size δ. Let $\mathbb{P}(\cdot)$ denote probability measure, and $\mathbb{E}[\cdot]$ the expectation with resp. to $\mathbb{P}$.

II. IAUP: Problem Formulation

Consider the following generalized information acquisition and utilization problem (IAUP) describing a Markov chain with M possible states. Let $S(t)$ describe the state of the system with a Markovian dynamic and according to transition matrix P. A Bayesian decision maker seeks to utilize the system by (sequentially) acquire the state of the system and subsequently utilize it. More structured than a general partially observable Markov decision problem (POMDP), the underlying dynamic is assumed to be independent of the decision maker's decision and actions. However, in contrast to hidden Markov model (HMM) formulation, our decision maker has the ability to control the collected observations' "information content" and hence the manner the system is utilized. In particular, we assume there are K acquisition and L utilization actions which can be employed. Information acquisition a costs C_a independently of the state of the system and provides an observation sample Z whose conditional distribution given the true state $S(t) = i$ is fixed and known, while utilization action u, provides a reward depending on how well u is "matched" to the true state of the system $S(t) = i$. Mathematically, the problem is described using the following setup:

Problem-IAUP

1) State $S_t \in \{1, 2, \ldots, M\}$ denotes the current state (out of M possible ones) of a Markov chain with transition matrix Q.
2) $\mathcal{Z}$ is the *observation space*. Observation $Z_t \in \mathcal{Z}$ denotes the observed sample at time t.
3) $\mathcal{A}$ is the *acquisition space* and is assumed to be finite with $|\mathcal{A}| = K < \infty$.
   - For all $a \in \mathcal{A}$, $s \in [M]$, and $z \in \mathcal{Z}$, the observation kernel $q_s^a(z)$ is the probability of observing $Z_t = z$ when action a has been taken and the current state of the process is s (independent of time and all other random variables).
   - Function $c : \mathcal{A} \to \mathbb{R}^+$, is the *one-step cost function associated with acquisition action a.*
   - It is often useful to allow for a special acquisition action $\star$ with zero acquisition cost under which no observation is collected (or equivalently the collected observation is independent of the current state). Exerting this action at time t, i.e. $a_t = \star$ is equivalent to skipping the sampling process.
4) $\mathcal{U}$ is the *utilization space* and is assumed to be finite with $|\mathcal{U}| = L < \infty$.
   - Functional $r : \mathcal{U} \times \{1, 2, \ldots, M\} \to \mathbb{R}^+$ describes the *one-step reward* $r(u,i)$ associated with taking utilization action u at state i.
5) Let $\mathcal{C} = \mathcal{A} \times \mathcal{U}$ be the set of allowable acquisition and utilization pairs.

Given a Bayesian prior ρ on the initial distribution of X_1, the above problem is a partially observable Markov decision problem (POMDP) where the state transition matrix is Q and the observations are noisy. The objective here is to find the optimal sequence of acquisition and utilization actions, $(A_t, U_t) \in \mathcal{C}$ in order to maximize the expected total weighted reward minus cost over a finite horizon:

$$J_T(\boldsymbol{\rho}) := \mathbb{E}\left[\sum_{t=1}^{T} r(U_t, S_t) - \lambda \sum_{t=1}^{T} c(A_t)\right]. \qquad (1)$$

or a discounted infinite horizon:

$$J_\beta(\boldsymbol{\rho}) := \mathbb{E}\left[\sum_{t=1}^{\infty} \beta^t r(U_t, S_t) - \lambda \sum_{t=1}^{\infty} \beta^t c(A_t)\right]. \qquad (2)$$

A slightly less stringent objective is that of minimizing the expected (long-term) average criterion:

$$J(\boldsymbol{\rho}) := \lim_{T\to\infty} \frac{1}{T} \mathbb{E}\left[\sum_{t=1}^{T} r(U_t, S_t) - \lambda \sum_{t=1}^{T} c(A_t)\right]. \qquad (3)$$

III. IAUP: Analysis

In this section, we provide two sets of structural results. In Subsection III-A, we show the equivalence between our general formulation of information aquisition and utilization problems and that of a generalized tacking. In Subsection III-B, we provide the partially observable Markov decision problem (POMDP) representation of the problem and the corresponding dynamic programs associated with the above cost criteria.

A. Equivalence to Generalized Tracking

Note that the posterior distribution of the state is strongly dependent on the acquisition action a through the Bayes' update but is not directly affected by the utilization action. Also the Markovian transition (prediction step) is not affected by either acquisition or utilization actions since it only incorporates the the transition matrix of the underlying Markov chain. This separated impact of acquisition and utilization actions, we will argue in this subsection, results in further simplification and analytic characterization of the solution to special classes of IAUP. In particular, we will show next that the problem of information acquisition and utilization is nothing but an active tracking problem:

Problem-GT

1) State $S_t \in \{1, 2, \ldots, M\}$ denotes the current state (out of M possible ones) of a Markov chain with transition matrix Q.
2) $\mathcal{Z}$ is the *observation space*. Observation $Z_t \in \mathcal{Z}$ denotes the observed sample at time t.
3) $\mathcal{A}$ is the *acquisition space* and is assumed to be finite with $|\mathcal{A}| = K < \infty$.
   - For all $a \in \mathcal{A}$, $s \in [M]$, and $z \in \mathcal{Z}$, the observation kernel $q_s^a(z)$ is the probability of observing $Z_t = z$ when action a has been taken and the current state of the process is s (independent of time and all other random variables).
   - Function $c : \mathcal{A} \to \mathbb{R}^+$, is the *one-step cost function associated with acquisition action* a.
   - It is often useful to allow for a special acquisition action $\star$ with zero acquisition cost under which no observation is collected (or equivalently the collected observation is independent of the current state). Exerting this action at time t, i.e. $a_t = \star$ is equivalent to skipping the sampling process.
4) $\mathcal{U}$ is the *declartion space* and is assumed to be finite with $|\mathcal{U}| = L < \infty$.
   - Let functional $\Delta : \mathcal{U} \times \{1, 2, \ldots, M\} \to \mathbb{R}^+$ describe the *distortion cost* $\Delta(i, u)$ associated with representing state i with element u.

Remark 1. The classic problem of active filtering/tracking is a special case of **Problem-GT**: let utilization action $U_t := \hat{S}_t \in \mathcal{U} = \mathcal{S}$, be the detection/tracking of the Markov state, via noisy sensing at the detection cost $r(\hat{S}_t, S_t) = \begin{cases} 1 & \text{if } \hat{S}_t = S_t \\ 0 & \text{if } \hat{S}_t \neq S_t \end{cases}$ plus the unit flat acquisition/sampling cost $c(A_t) = \begin{cases} 0 & \text{if } A_t = \star \\ 1 & \text{otherwise} \end{cases}$.

Furthermore, we note that any **Problem-IAUP** can be written as a **Problem-GT**. To see this, define $r^*(i) := \max_u r(u, i)$, for $i = 1, 2, \ldots, M$. Now the distortion function $\Delta(i, u) := r^*(i) - r(u, i)$ has the interpretation of the mismatch cost between state i and element u (in particular, there exists $u^*(i) = \max_u r(u, i)$ for which $\Delta(i, u^*(i)) = 0$). Now consider the objective of minimizing the overall cost over a finite horizon:

$$\tilde{J}_T(\boldsymbol{\rho}) := \mathbb{E}\left[\sum_{t=1}^{T} \Delta(U_t, S_t) + \lambda \sum_{t=1}^{T} c(A_t)\right]. \quad (4)$$

or a discounted infinite horizon:

$$\tilde{J}_\beta(\boldsymbol{\rho}) := \mathbb{E}\left[\sum_{t=1}^{\infty} \beta^t \Delta(U_t, S_t) + \lambda \sum_{t=1}^{\infty} \beta^t c(A_t)\right] \quad (5)$$

and the expected (long-term) average criterion:

$$\tilde{J}(\boldsymbol{\rho}) := \lim_{T\to\infty} \frac{1}{T} \mathbb{E}\left[\sum_{t=1}^{T} \Delta(U_t, S_t) + \lambda \sum_{t=1}^{T} c(A_t)\right]. \quad (6)$$

It is straightforward to show that:

$$J_T(\boldsymbol{\rho}) = \bar{R}_T(\rho) - \tilde{J}_T(\boldsymbol{\rho}),$$
$$J_\beta(\boldsymbol{\rho}) = \bar{R}_\beta(\rho) - \tilde{J}_\beta(\boldsymbol{\rho}),$$
$$J(\boldsymbol{\rho}) = \bar{R}(\rho) - \tilde{J}(\boldsymbol{\rho})$$

where

$$\bar{R}_T(\boldsymbol{\rho}) = \sum_{t=1}^{T} < \boldsymbol{\rho} Q^{t-1}, \boldsymbol{r}^* >,$$
$$\bar{R}_\beta(\boldsymbol{\rho}) = \sum_{t=1}^{\infty} \beta^t < \boldsymbol{\rho} Q^{t-1}, \boldsymbol{r}^* >,$$
$$\bar{R}(\boldsymbol{\rho}) = \frac{1}{T} \sum_{t=1}^{T} < \boldsymbol{\rho} Q^{t-1}, \boldsymbol{r}^* >,$$

i.e. the first term after each equality is equal to the maximum expected utilization reward if the observations were perfect, and hence is independent of the choice of actions chosen. This proves that the original IAUP can be written in form of a generalized active state tracking problem in noise, where the acquisition actions are sequentially selected to minimize the overall distoriton-plus-acquisition-cost.

Remark 2. The problem real-time joint Markov source–channel coding over a discrete memoryless channel with feedback [5] can also be cast as a special case of **Problem-GT** where the space of acquisition actions is the set of allowable encoding functions and the space of utilization actions coincides with the state space of the Markov source.

B. Dynamic Programming Equations

Given a Bayesian prior ρ on the initial distribution of X_1, our POMDP problem is equivalent to an MDP whose information state (sufficient statistics) at time t is the belief vector $\boldsymbol{\rho}(t) = [\rho_1(t), \ldots, \rho_M(t)]$ where

$$\rho_i(t) = \text{Prob}(X_t = i \,|Z_{1:t}),$$

and the information state space is defined as $\mathbb{P}(\Omega) = \{\boldsymbol{\rho} \in [0, 1]^M : \sum_{i=1}^{M} \rho_i = 1\}$.

Let $\Phi^a(\boldsymbol{\rho}, z)$ denote the posterior belief state, given current belief (posterior), the acquisition action a, and channel output z. This means that for all $i \in \Omega$, posterior belief

$$\Phi^a(\boldsymbol{\rho}(t), z)(i) := \frac{\rho_i^+(t) q_a(z|i)}{\sum_{j=1}^{M} \rho_j^+(t) q_a(z|x)},$$

where

$$\rho_i^+(t) := \mathbb{P}(X_{t+1} = i|Z_{1:t}) = \sum_{j\in\Omega} \rho_j(t) Q_{ji}$$

is the *pre-transmission one step predictor* of the state.

Furthermore, for an acquisition action a, define the corresponding Markov operator $\mathbb{T}^a : \mathcal{M}(\mathbb{P}(\Omega)) \to \mathcal{M}(\mathbb{P}(\Omega))$ as

$$(\mathbb{T}^a g)(\boldsymbol{\rho}) := \mathbb{E}\left[g\left(\Phi^a(\boldsymbol{\rho}Q, Z)\right)\right],$$

where the expectation is taken with respect to conditional distribution of Z given action a and prior $\boldsymbol{\rho}$. In addition, for utilization action u and $\boldsymbol{\rho} \in \mathbb{P}(\Omega)$, to allow for ease of notation, let us denote by $\Delta(\boldsymbol{\rho}, u) = \sum_i \rho_i(t)\Delta(i, u)$ the expected utilization reward and by $\Delta^*(\boldsymbol{\rho}) = \min_{u\in\mathcal{U}} \Delta(\boldsymbol{\rho}, u)$.

Fact 1 (Propositions 3.1 in [6])**.** *Define recursively the functions:*

$$V_T(\boldsymbol{\rho}) = \min_{u\in\mathcal{U}} \Delta(\boldsymbol{\rho}, u) \tag{7}$$

$$\begin{aligned} V_t(\boldsymbol{\rho}) &= \min_{(a,u)\in\mathcal{C}} \Delta(\boldsymbol{\rho}, u) + \lambda c(a) + (\mathbb{T}^a V_{t+1})(\boldsymbol{\rho}) \\ &= \min_{u\in\mathcal{U}} \Delta(\boldsymbol{\rho}, u) + \min_{a\in\mathcal{A}} \left[\lambda c(a) + (\mathbb{T}^a V_{t+1})(\boldsymbol{\rho})\right] \\ &= \Delta^*(\boldsymbol{\rho}) + \min_{a\in\mathcal{A}} \left[\lambda c(a) + (\mathbb{T}^a V_{t+1})(\boldsymbol{\rho})\right]. \end{aligned} \tag{8}$$

Then $V_1(\boldsymbol{\rho}_1)$, known as the optimal value function at $t = 1$, is equal to the minimum cost $\tilde{J}_T(\boldsymbol{\rho}_1)$ in (4) with the prior belief $\boldsymbol{\rho}_1$.

Important consequences of the above set of dynamic programming equations are as follows.

1) The minimizers of (7) and (9) constitute deterministic optimal Markov policy (selecting acquisition and declaration actions deterministic functions of the sufficient statistic $\boldsymbol{\rho}$).
2) The optimal acquisition and utilization can be selected in a separated fashion.
3) The optimal utilization is the Bayes' risk minimizer which, at any given belief vector $\boldsymbol{\rho}$, minimizes the expected mismatch between the utilization action and the (hidden) state.
4) If $V_{t+1}(\boldsymbol{\rho}Q) \le c(a) + (\mathbb{T}^a V_{t+1})(\boldsymbol{\rho})$ for all $a \subset \mathcal{A} - \{\star\}$, then it is optimal to skip the next sample (take utilization action $\star$). The result can be generalized for skipping the next τ transmissions.

Similar results can be obtained for $\tilde{J}_\beta(\boldsymbol{\rho})$.

Fact 2. *Consider the solution to the following fixed point equation:*

$$V^\beta(\boldsymbol{\rho}) = \Delta^*(\boldsymbol{\rho}) + \min_{a\in\mathcal{A}} \left[\lambda c(a) + \beta(\mathbb{T}^a V^\beta)(\boldsymbol{\rho})\right]. \tag{9}$$

Then $V^\beta(\boldsymbol{\rho}_1)$, known as the optimal value function, is equal to the minimum cost $\tilde{J}_\beta(\boldsymbol{\rho}_1)$ with the prior belief $\boldsymbol{\rho}_1$.

The next theorem establishes the dynamic programming optimality equation for the expected average cost criterion:

Fact 3. *If there exists scalar V^* and bounded function $W \in \mathcal{M}(\mathbb{P}\Omega)$ such that*

$$V^* + W(\boldsymbol{\rho}) = \Delta^*(\boldsymbol{\rho}) + \min_a \left[\lambda c(a) + (\mathbb{T}^a W)(\boldsymbol{\rho})\right], \tag{10}$$

then V^* is equal to the minimum cost $\tilde{J}(\boldsymbol{\rho}_1)$ in (6) (which becomes independent of initial prior).

Similarly, the above dynamic programming result (and the corresponding optimality equation) has the following simple consequences:

1) The minimizer of (10) constitutes a deterministic *stationary* optimal Markov policy (selecting acquisition and declaration actions as deterministic and *time invariant* functions of the sufficient statistic $\boldsymbol{\rho}$):
   - The optimal utilization is the Bayes' risk minimizer which, at any given belief vector $\boldsymbol{\rho}$, maximizes the expected reward (minimizing the expected risk).
   - The optimal acquisition is given as :

$$\arg\max_a \{W(\boldsymbol{\rho}) - (\mathbb{T}^a W)(\boldsymbol{\rho}) - \lambda c(a)\} \tag{11}$$

2) An important question related to Fact 3 is if and when V^* can be approximated by considering the finite horizon problem of (7)-(9) as T gets large [7]. Similarly, another important question relies on the vanishing discount technique with an infinite horizon discounted cost criterion. Such approximation results, if proved, would enable computational solutions via value/policy iteration.

So far these results depend on the existence of V^* and function W satisfying (10). An important question is the sufficient and necessary conditions for the existence of V^* and function W satisfying (10). The following proposition, whose proof follows from [7], provides a partial answer.

Theorem 1. *Assume that the state transition matrix Q is aperiodic and irreducible. In addition, assume that $P^a(\cdot|X = x)$ is absolutely continuous with respect to $P^a(\cdot|X = x')$, for all $x, x' \in \mathcal{X}$. Under these conditions, there exist V^* and bounded function W satisfying (10).*

Intuitively, the optimal solution to this optimization problem requires sufficiently accurate characterization of functions $V_1, V_2, \ldots, V_T$ as well as W. Such a characterization, however, depends on the designer's ability to identify the optimal trade-off between the cost of information acquisition versus the corresponding utilization rewards, as well as a trade-off between gathering information now versus shaping the distributions in the hopes of high informational return. A possible solution is to compute these functions numerically using the value iteration technique [8, Chapter 9.5] or to estimate it using basis functions. While there are some theoretical guarantees regarding the convergence to the optimal value function [8, Proposition 9.17], the value iteration quickly becomes intractable as the number of hypotheses grows; in contrast, the latter method of basis function approximation is easier to implement but provides no convergence guarantee. In lieu of numerical approximation of or derivation of a closed-form for the optimal value, we introduce alternative heuristics. Before we discuss these heuristics, however, we consider the very special case of stochastic degradation. In particular, consider

Definition 1 (Blackwell Ordering [9])**.** Given two groups of conditional probability densities $q^a = \{q_i^a\}_{i\in\Omega}$ and $q^b = \{q_i^b\}_{i\in\Omega}$ (on space $\mathcal{Z}$), we say that q^b is *less informative than* q^a ($q^b \leq_B q^a$) if there exists a *stochastic transformation* W from $\mathcal{Z}$ to $\mathcal{Z}$ such that[1]

$$q_i^b(z) = \int q_i^a(y)W(y;z)dy \quad \forall i \in \Omega. \tag{12}$$

The following fact is an important outcome of Blackwell ordering.

Fact 4 (see [10] ch. 14.17)**.** *Let* $q^a = \{q_i^a\}_{i\in\Omega}$ *and* $q^b = \{q_i^b\}_{i\in\Omega}$ *be two groups of observation kernels. If* $q^b \leq_B q^a$*, then* $(\mathbb{T}^a g)(\boldsymbol{\rho}) \leq (\mathbb{T}^b g)(\boldsymbol{\rho})$ *for all* $\boldsymbol{\rho} \in \mathbb{P}(\Omega)$ *and for any concave function* $g : \mathbb{P}(\Omega) \to \mathbb{R}$.

The following lemma provides the technical condition to ensure relevance of the above Fact to our DP formulation.

Lemma 1. *Functions* V_1, V_2, $\ldots$, V_T *as well as* W *are convex in belief state* $\boldsymbol{\rho}$.

C. Information Utility

As discussed earlier, in lieu of numerical approximation of or derivation of a closed-form for V^*, in this subsection, we introduce alternative notions of information maximizing which gives rise to simple deterministic and Markov heuristic policies. Note that characterization of the optimal policy in the previous subsection has the following interpretation associated with the following notion of information utility.

Definition 2. Associated with a concave functional $h : \mathbb{P}(\Theta) \to \mathbb{R}+$, the *information utility* of action a at information state $\boldsymbol{\rho}$ is defined as $\mathcal{IU}(a, \boldsymbol{\rho}, h) := h(\boldsymbol{\rho}) - (\mathbb{T}^a h)(\boldsymbol{\rho})$.

Together with (11), this results in the following optimal deterministic choice of acquisition action:

$$\arg\max_a \{\mathcal{IU}(e, \boldsymbol{\rho}, W) - \lambda c(a)\}$$

maximizing the information utility minus cost at any given belief $\boldsymbol{\rho}$. This suggests that by choosing an appropriate notion of information utility, any IAUP can be reduced to a sequence of one-shot problems in each of which an optimal sensing action is selected deterministically so as to provide the highest amount of information. Next, we give the definition of *Jensen–Shannon (JS) divergence*, and we introduce *Extrinsic Jensen–Shannon (EJS) divergence*. In Section III-D, we show that JS and EJS divergences are equal to information relative to the Shannon entropy and average log-likelihood function, respectively.

1) Symmetric Divergences: We first recall some well known divergences. The *Kullback–Leibler (KL) divergence* between two probability distributions P_Z and P'_Z over a finite set $\mathcal{Z}$ is defined as $D(P_Z\|P'_Z) := \sum_{z\in\mathcal{Z}} P_Z(z)\log\frac{P_Z(z)}{P'_Z(z)}$ with the convention $0\log\frac{a}{0} = 0$ and $b\log\frac{b}{0} = \infty$ for $a, b \in [0,1]$ with $b \neq 0$. The KL divergence satisfies the following lemma.

Lemma 2. *For any two distributions* P *and* Q *on a set* $\mathcal{Z}$ *and* $\alpha \in [0,1]$*,* $D(P\|\alpha P + (1-\alpha)Q)$ *is decreasing in* α.

The KL divergence is *not* symmetric, i.e., in general

$$D(P_Z\|P'_Z) \neq D(P'_Z\|P_Z).$$

The *J divergence* [11] and *L divergence* [12] symmetrize the KL divergence:

$$J(P_1, P_2) := D(P_1\|P_2) + D(P_2\|P_1), \tag{13}$$

$$L(P_1, P_2) := D\Big(P_1\|\frac{1}{2}P_1 + \frac{1}{2}P_2\Big) + D\Big(P_2\|\frac{1}{2}P_1 + \frac{1}{2}P_2\Big). \tag{14}$$

The L divergence can be related to the *Jensen difference* with respect to the Shannon entropy function [13]:

$$\frac{1}{2}L(P_1, P_2) = H\left(\frac{1}{2}P_1 + \frac{1}{2}P_2\right) - \left(\frac{1}{2}H(P_1) + \frac{1}{2}H(P_2)\right). \tag{15}$$

The *Jensen–Shannon (JS) divergence* [13], [12] is defined as an M-dimensional generalization of the L divergence. Given M distributions $P_1, P_2 \ldots, P_M$ over a set $\mathcal{Z}$ and a vector of a priori weights $\boldsymbol{\rho} = [\rho_1, \rho_2, \ldots, \rho_M]$, where $\boldsymbol{\rho} \in [0,1]^M$ and $\sum_{i=1}^M \rho_i = 1$, the JS divergence is defined as [13], [12]:

$$\begin{aligned} JS(\boldsymbol{\rho}; P_1, \ldots, P_M) &:= \sum_{i=1}^M \rho_i D\Big(P_i\|\sum_{j=1}^M \rho_j P_j\Big) \\ &= H\Big(\sum_{i=1}^M \rho_i P_i\Big) - \sum_{i=1}^M \rho_i H(P_i). \end{aligned} \tag{16}$$

Let θ be a random variable that takes values in $\{1, 2, \ldots, M\}$ and has probability mass function $\boldsymbol{\rho}$ and $Z \sim P_\theta$ (which implies that $P(Z = z) = \sum_{i=1}^M \rho_i P_i(z)$). From (16),

$$JS(\boldsymbol{\rho}; P_1, \ldots, P_M) = H(Z) - H(Z|\theta) = I(\theta; Z), \tag{17}$$

where $I(\theta; Z)$ is the mutual information between θ and Z.

From (17) and the fact that $I(\theta; Z) = H(\theta) - H(\theta|Z)$, the JS divergence can also be expressed as:

$$JS(\boldsymbol{\rho}; P_1, \ldots, P_M) \tag{18}$$

$$= H(\boldsymbol{\rho}) - \sum_{z\in\mathcal{Z}} P_{\boldsymbol{\rho}}(z) H\Big(\Big[\frac{\rho_1 P_1(z)}{P_{\boldsymbol{\rho}}(z)}, \ldots, \frac{\rho_M P_M(z)}{P_{\boldsymbol{\rho}}(z)}\Big]\Big), \tag{19}$$

where $P_{\boldsymbol{\rho}} = \sum_{i=1}^M \rho_i P_i$.

[1] Function $W : \mathcal{Z} \times \mathcal{Z} \to \mathbb{R}_+$ is called a *stochastic transformation* from $\mathcal{Z}$ to $\mathcal{Z}$ if it satisfies $\int_{\mathcal{Z}} W(y;z)dz = 1$ for all $y \in \mathcal{Z}$ and $\int_{\mathcal{Z}} W(y;z)dy < \infty$ for all $z \in \mathcal{Z}$.

2) A New Divergence: Extrinsic Jensen–Shannon (EJS) Divergence: We introduce the *Extrinsic Jensen–Shannon (EJS) divergence* as an M-dimensional generalization of J divergence as

$$EJS(\boldsymbol{\rho}; P_1, \ldots, P_M) := \sum_{i=1}^{M} \rho_i D\Big(P_i \| \sum_{j \neq i} \frac{\rho_j}{1-\rho_i} P_j\Big), \tag{20a}$$

when $\rho_i < 1$ for all $i \in \{1, \ldots, M\}$, and as

$$EJS(\boldsymbol{\rho}; P_1, \ldots, P_M) := \max_{j \neq i} D(P_i \| P_j) \tag{20b}$$

when $\rho_i = 1$ for some $i \in \{1, \ldots, M\}$.

Let $U(\cdot)$ denote the average log-likelihood function:

$$U(\boldsymbol{\rho}) := \sum_{i=1}^{M} \rho_i \log \frac{1-\rho_i}{\rho_i}. \tag{21}$$

Lemma 3 (Properties of EJS Divergence)**.** *The EJS divergence defined in* (20) *satisfies the following three properties.*

1) *It is lower bounded by the JS divergence:*

$$EJS(\boldsymbol{\rho}; P_1, \ldots, P_M) \geq JS(\boldsymbol{\rho}; P_1, \ldots, P_M). \tag{22}$$

2) *It can be expressed as*

$$EJS(\boldsymbol{\rho}; P_1, \ldots, P_M) \tag{23}$$

$$= U(\boldsymbol{\rho}) - \sum_{z \in \mathcal{Z}} P_{\boldsymbol{\rho}}(z) U\Big(\Big[\frac{\rho_1 P_1(z)}{P_{\boldsymbol{\rho}}(z)}, \ldots, \frac{\rho_M P_M(z)}{P_{\boldsymbol{\rho}}(z)}\Big]\Big). \tag{24}$$

3) *It is convex in the distributions* $P_1, \ldots, P_M$.

Remark 3. The EJS divergence defined in this section is not the unique generalization of the J divergence. There exist other M-dimensional generalizations of the J divergence such as $\sum_{i=1}^{M} \rho_i \sum_{j=1}^{M} \rho_j J(P_i, P_j)$ which was studied in [14]. However, as will be discussed in details later, properties of EJS such as the one provided by (23) above makes it a proper notion of information for our applications of interest.

D. Heuristic Policies: Maximizing (E)JS Divergence

In this section, we show that JS and EJS divergences are proper notions of information, and propose deterministic Markov policies based on greedy maximization of these divergences.

Given a belief vector $\boldsymbol{\rho} \in \mathbb{P}(\Omega)$ and a sensing action $a \in \mathcal{A}$, we use the notations

$$JS(\boldsymbol{\rho}, a) := JS(\boldsymbol{\rho}; q_1^a, q_2^a, \ldots, q_M^a), \tag{25}$$

$$EJS(\boldsymbol{\rho}, a) := EJS(\boldsymbol{\rho}; q_1^a, q_2^a, \ldots, q_M^a). \tag{26}$$

Definition 2 together with properties (16) and (23) shows that the JS and EJS divergences are nothing but the information relative to the entropy and average log-likelihood functions, respectively, i.e.,

$$JS(\boldsymbol{\rho}, a) \tag{27}$$

$$= \mathcal{I}(\boldsymbol{\rho}, a, H), \tag{28}$$

$$EJS(\boldsymbol{\rho}, a) = \mathcal{I}(\boldsymbol{\rho}, a, U). \tag{29}$$

We also use the following notations to denote the amount of information a Markov stationary policy π obtains in a single step:

$$JS(\boldsymbol{\rho}, \pi) := \sum_{a \in \mathcal{A}} \pi(a|\boldsymbol{\rho}) JS(\boldsymbol{\rho}, a), \tag{30}$$

$$EJS(\boldsymbol{\rho}, \pi) := \sum_{a \in \mathcal{A}} \pi(a|\boldsymbol{\rho}) EJS(\boldsymbol{\rho}, a). \tag{31}$$

We are now ready to introduce our heuristic policies.

Policy π_{JS} is defined as follows: given belief state $\boldsymbol{\rho}(t)$ select $\arg\max_{a \in \mathcal{A}} JS(\boldsymbol{\rho}(t), a)$.

Similarly, policy π_{EJS} is defined as follows: given the belief state $\boldsymbol{\rho}(t)$, select $\arg\max_{a \in \mathcal{A}} EJS(\boldsymbol{\rho}(t), a)$.

Remark 4. Note that as the belief about one of the hypotheses, say ρ_i, approaches 1, $D(q_i^a \| \sum_{j=1}^{M} \rho_j q_j^a)$ converges to $D(q_i^a \| q_i^a) = 0$ for any action $a \in \mathcal{A}$; and consequently, independently of the observation kernels $q_1^a, q_2^a, \ldots, q_M^a$, divergence $JS(\boldsymbol{\rho}, a)$ approaches 0. In contrast, as ρ_i becomes large, $EJS(\boldsymbol{\rho}, a)$ approaches $D(q_i^a \| \sum_{j \neq i} \frac{\rho_j}{1-\rho_i} q_j^a)$ and hence, π_{EJS} selects action a such that $D(q_i^a \| \sum_{j \neq i} \frac{\rho_j}{1-\rho_i} q_j^a)$ is maximized, i.e., it selects an action that distinguishes H_i from the collection of alternate hypotheses the most. As we will see in the next section, these different philosophies result in significant performance difference.

1) Numerical Example: Consider an IAUP with the time-invariant and binary hidden state, whose nadditive 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 is such that it adds unequal noise to the hypotheses.

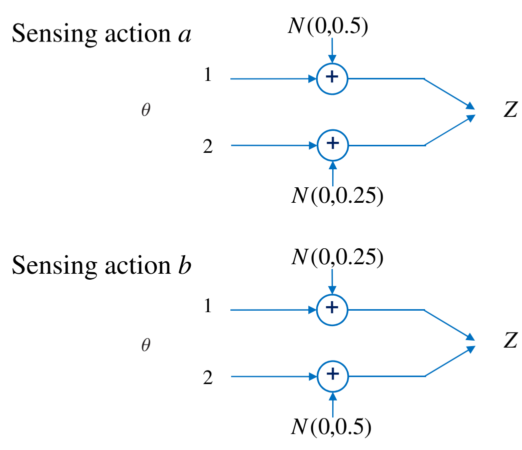

Fig. 1. Active binary hypothesis testing with additive Gaussian noise.

Figures 2 and 3 show respectively the JS and EJS divergence for all $\boldsymbol{\rho} \in \mathbb{P}(\Omega)$ and for sensing actions a and

b. It is clear from these figures that π_{JS} selects action a when $\rho_1 \geq 0.5$ and selects action b otherwise; while π_{EJS} does exactly the opposite. These figures illustrate the fact mentioned in Remark 4.

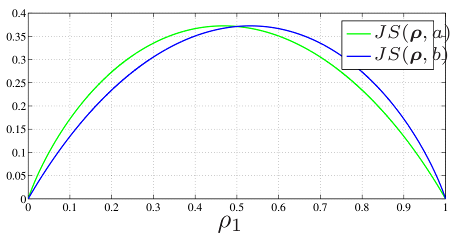

Fig. 2. JS divergence for sensing actions a and b.

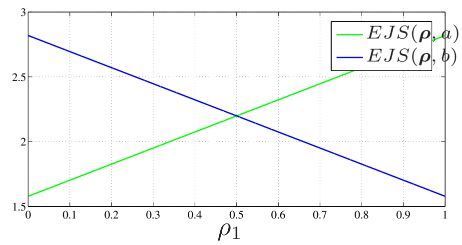

Fig. 3. EJS divergence for sensing actions a and b.

IV. Active Hypothesis Testing

A. Problem Formulation

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

Problem-AHT

1) Let H_i, $i \in \Omega = \{1, 2, \ldots, M\}$, denote M hypotheses of interest among which only one holds true. Let θ 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 with a given prior (belief) about θ, i.e., initially $P(\theta = i) = \rho_i(0) > 0$ for all $i \in \Omega$.
2) $\mathcal{A}$ is the set of all sensing actions and is assumed to be finite with $|\mathcal{A}| = K < \infty$.
3) $\mathcal{Z}$ 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 $\{q_i^a(\cdot)\}_{i\in\Omega, a\in\mathcal{A}}$ are known, and the observations are conditionally independent over time.
4) Let L, $L > 1$, denote the penalty for a wrong declaration, i.e., the penalty of selecting H_j, $j \neq i$, when H_i is true.[2]
5) Let τ be the (stopping) time at which the decision maker retires.
6) The objective is to find a stopping time τ, a sequence of sensing actions $A(0), A(1), \ldots, A(\tau - 1)$, and a declaration rule $d:\ \mathcal{A}^\tau \times \mathcal{Z}^\tau \to \Omega$ that collectively minimize the expected total cost
$$\mathbb{E}\left[\tau + L\mathbf{1}_{\{d(A^{\tau-1}, Z^{\tau-1}) \neq \theta\}}\right], \tag{32}$$
where $A^{\tau-1} = [A(0), \ldots, A(\tau-1)]$, $Z^{\tau-1} = [Z(0), \ldots, Z(\tau-1)]$, and the expectation is taken with respect to the initial belief on θ as well as the distributions of action sequence, observation sequence, and the stopping time.

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. This problem is a special case of IAUP problem where the state is fixed with a termination state (resulting in a shortest path version of the problems of cumulative cost studied earlier.)

We are interested in the characterization of this cost as a function of penalty L and the number of hypotheses M. It is easy to show that under the optimal selection rule, the probability of error approaches zero as L approaches infinity[3].

B. Analysis: Dynamic Programming

Note that at any given information state $\boldsymbol{\rho}$, taking sensing action $a \in \mathcal{A}$ followed by the optimal policy results in expected total cost $1 + (\mathbb{T}^a V^*)(\boldsymbol{\rho})$ where 1 denotes the one unit of time spent to take the sensing action and collect the corresponding observation sample, and $(\mathbb{T}^a V^*)(\boldsymbol{\rho})$ is the expected value of V^* on the space of posterior beliefs; while declaration j results in expected cost $(1-\rho_j)L$ where $(1-\rho_j)$ is the probability that hypothesis H_j is not true, and L is the penalty of making a wrong declaration. This intuition, while relying on the compactness of $\mathbb{P}(\Omega)$ to treat various measurability issues, can be formalized in the following dynamic programming equation.

Theorem 2. *The optimal value function $V^* : \mathbb{P}(\Omega) \to \mathbb{R}_+$ is the unique solution to the following fixed point DP equation:*
$$V^*(\boldsymbol{\rho}) = \min\left\{1 + \min_{a\in\mathcal{A}} (\mathbb{T}^a V^*)(\boldsymbol{\rho}), \min_{j\in\Omega} (1-\rho_j)L\right\}. \tag{33}$$

Now this DP characterization can be used to find the following lower bounds.

C. Analysis: Lower Bounds on the Optimal Performance

We rely on the following Assumptions in order to obtained lower bound on the optimal performance.

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

Assumption 2. *There exists $\xi_M < \infty$ such that*
$$\max_{i,j\in\Omega_M} \max_{a\in\mathcal{A}_M} \sup_{z\in\mathcal{Z}} \log \frac{q_i^a(z)}{q_j^a(z)} \leq \xi_M.$$

[2]In general, we can define a loss matrix $[L_{ij}]_{i,j\in\Omega}$, where L_{ij} denotes the penalty (loss) of selecting H_j when H_i is true.

[3]. In particular, it can be shown that 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 probability of error [2].

Assumption 3 ensures the possibility of discrimination between any two hypotheses, hence ensuring that the problem of active hypothesis testing has a meaningful solution. Assumption 5 implies that no two hypotheses are fully distinguishable using a single observation. Assumption 5 is a technical one which enables our non-asymptotic characterizations of the upper and lower bounds. In Subsection IV-E, we discuss the consequence of weakening this assumption in detail.

If there exists a functional $V : \mathbb{P}(\Omega_M) \to \mathbb{R}_+$ such that for all belief vectors $\boldsymbol{\rho} \in \mathbb{P}(\Omega_M)$,

$$V(\boldsymbol{\rho}) \leq \min\{1 + \min_{a\in\mathcal{A}_M} (\mathbb{T}^a V)(\boldsymbol{\rho}), \min_{j\in\Omega_M} (1-\rho_j)L\},$$

then $V^*(\boldsymbol{\rho}) \geq V(\boldsymbol{\rho})$ for all $\boldsymbol{\rho} \in \mathbb{P}(\Omega_M)$. In this section, we use the above fact to find three lower bounds for V^*. These lower bounds are non-asymptotic and complementary for various values of the parameters of the problem.

Theorem 3. *Under Assumption 3 and for $L > 1$ and $\boldsymbol{\rho} \in \mathbb{P}_L(\Omega_M)$,*

$$V^*(\boldsymbol{\rho}) \geq \underline{V_1}(\boldsymbol{\rho}) := \left[\sum_{i=1}^{M} \rho_i \max_{j\neq i} \frac{\log\frac{1-L^{-1}}{L^{-1}} - \log\frac{\rho_i}{\rho_j}}{\max\limits_{a\in\mathcal{A}_M} D(q_i^a \| q_j^a)} - K_1'\right]^+,$$

where K_1' is a constant independent of L.

Let define

$$D_{\max}(M) := \max_{i,j\in\Omega_M} \max_{a\in\mathcal{A}_M} D(q_i^a \| q_j^a), \tag{34}$$

$$I_{\max}(M) := \max_{a\in\mathcal{A}_M} \max_{\hat{\boldsymbol{\rho}}\in\mathbb{P}(\Omega_M)} JS(\hat{\boldsymbol{\rho}}, a). \tag{35}$$

Next we provide another lower bound which is more appropriate for large values of M or small values of $I_{\max}(M)$.

Theorem 4. *Under Assumption 3 and for $L > 1$ and $\boldsymbol{\rho} \in \mathbb{P}_L(\Omega_M)$,*

$$V^*(\boldsymbol{\rho}) \geq \underline{V_2}(\boldsymbol{\rho}) := \left[\frac{H(\boldsymbol{\rho}) - H([\alpha(L,M), 1-\alpha(L,M)])}{I_{\max}(M)} - \frac{\alpha(L,M)\log(M-1)}{I_{\max}(M)} + \alpha(L,M)L\right]^+,$$

where $\alpha(L,M) := \frac{M-1}{M-1+2^{LI_{\max}(M)}}$.

Remark 5. The lower bounds in Theorems 3 and 4 can be explained by the following intuition: For any uncertainty function $V : \mathbb{P}(\Omega_M) \to \mathbb{R}$, the number of samples required to reduce the uncertainty down to a target level V_{target} has to be at least $\frac{V(\boldsymbol{\rho}(0)) - V_{\text{target}}}{\Delta_{\max}(V)}$, where $\Delta_{\max}(V)$ is the maximum amount of reduction in V associated with a single sample, i.e., $\Delta_{\max}(V) = \max\limits_{a\in\mathcal{A}_M} \max\limits_{\boldsymbol{\rho}\in\mathbb{P}(\Omega_M)} \mathcal{I}(\boldsymbol{\rho}, a, V)$. The lower bound in Theorem 3 is associated with such a lower bound when taking V to be the log-likelihood function, while the lower bound in Theorem 4 is associated with setting V to be the Shannon entropy.

Theorem 4 can be used to show that when $L < \frac{\log M}{I_{\max}(M)}$, the problem of active hypothesis testing will have a trivial solution. The precise statement is given by the following corollary.

Corollary 1. *Let $L < \frac{\log M}{I_{\max}(M)}$, and suppose the decision maker has a uniform prior belief about the hypotheses. For sufficiently large M, the optimal policy randomly guesses the true hypothesis without collecting any observation, hence,* Pe*, the probability of making a wrong declaration, approaches $1 - \frac{1}{M}$.*

Next theorem combines the above lower bounds and is appropriate when L and M are both large.

Theorem 5. *Under Assumptions 3 and 5, and for $L > \max\{1, \frac{\log M}{I_{\max}(M)}\}$ and arbitrary $\delta \in (0, 0.5]$,*

$$V^*(\boldsymbol{\rho}) \geq \underline{V_3}(\boldsymbol{\rho}) := \left[\frac{H(\boldsymbol{\rho}) - H([\delta, 1-\delta]) - \delta\log(M-1)}{I_{\max}(M)} + \frac{\log\frac{1-L^{-1}}{L^{-1}} - \log\frac{1-\delta}{\delta} - \xi_M}{D_{\max}(M)} \mathbf{1}_{\{\max\limits_{i\in\Omega_M} \rho_i \leq 1-\delta\}} - K_3'\right]^+, \tag{36}$$

where K_3' is a constant independent of δ and L.[4]

The next proposition generalizes this lower bound and is provided as a benchmark for comparison with the lower bounds derived using dynamic programming techniques.

Proposition 1. *Under Assumptions 3 and 5, and for $L > 1$, $\boldsymbol{\rho} \in \mathbb{P}_L(\Omega_M)$, and arbitrary $\delta \in (0,1)$,*

$$V^*(\boldsymbol{\rho}) \geq \left[\sum_{i=1}^{M} \rho_i \frac{\left[(1-\delta)\log\frac{L}{K'\log 2L} - \max\limits_{j\neq i}\log\frac{\rho_i}{\rho_j}\right]^+}{\tilde{D}_i(M) + \delta} \left(1 - \frac{2M(\frac{K'\log 2L}{L})^\delta}{\rho_i}\right) - \frac{M\xi_M^2}{\delta^2}\right]^+, \tag{37}$$

where K' is a constant independent of δ and L.

D. Analysis: Upper Bounds and Achievability

In this section, we analyze the performance of the proposed heuristics, and provide upper bounds on the expected total cost achieved by these policies. Let $V_\pi(\boldsymbol{\rho}) := \mathbb{E}_\pi[\tau] + L\mathrm{Pe}_\pi$ denote the expected total cost (32) achieved by policy π given that the initial belief is $\boldsymbol{\rho}$.

Theorem 6. *Consider a policy π that selects the retire-declare action at*

$$\tilde{\tau}_{1/L} = \min\{t : \max_{i\in\Omega_M} \rho_i(t) \geq 1 - L^{-1}\}. \tag{38}$$

[4]It can be shown that K_3' can be selected independent of M as well if $\sup_M \xi_M < \infty$.

Suppose policy π at each time $t = 0, \dots, \tilde{\tau}_{1/L} - 1$ and given the posterior vector $\boldsymbol{\rho}(t)$, selects sensing actions in a way that $EJS(\boldsymbol{\rho}(t), \pi) \geq \alpha$ for some $\alpha > 0$. Under Assumptions 3 and 5, and for $L > 1$ and $\boldsymbol{\rho} \in \mathbb{P}_L(\Omega_M)$,

$$V_\pi(\boldsymbol{\rho}) \leq \overline{V}_\alpha(\boldsymbol{\rho}) \tag{39}$$

$$:= \frac{H(\boldsymbol{\rho}) + \max\{\log\log M, \log L\} + 2^{\xi_M + 2}}{\alpha} + 1. \tag{40}$$

Furthermore, if there exist positive values α and β such that at each time $t < \tilde{\tau}_{1/L}$ and given the posterior vector $\boldsymbol{\rho}(t)$,

$$EJS(\boldsymbol{\rho}(t), \pi) \geq \begin{cases} \alpha & \text{if } \max_{i \in \Omega_M} \rho_i(t) < \tilde{\rho} \\ \beta & \text{otherwise} \end{cases}, \tag{41}$$

where

$$\tilde{\rho} := 1 - \frac{1}{1 + \max\{\log M, \log L\}}, \tag{42}$$

then the following bound is obtained

$$V_\pi(\boldsymbol{\rho}) \leq \overline{V}_{\alpha\beta}(\boldsymbol{\rho}) \tag{43}$$

$$:= \frac{H(\boldsymbol{\rho}) + \max\{\log\log M, \log\log L\}}{\alpha} + \frac{\log L}{\beta} \tag{44}$$

$$+ \frac{3 \times 2^{2\xi_M + 4}}{\alpha\beta} + 1. \tag{45}$$

Recall from Section III-D that policy π_{EJS} selects the retire-declare action at $\tilde{\tau}_{1/L}$ and selects sensing actions in a way to maximize the EJS divergence, i.e.,

$$EJS(\boldsymbol{\rho}, \pi_{EJS}) = \max_{a \in \mathcal{A}_M} EJS(\boldsymbol{\rho}, a). \tag{46}$$

Theorem 6 together with (46) yields the following:

Corollary 2. *Suppose there exist positive values α and β such that $\forall \boldsymbol{\rho} \in \mathbb{P}_L(\Omega_M)$,*

$$\max_{a \in \mathcal{A}_M} EJS(\boldsymbol{\rho}, a) \geq \begin{cases} \alpha & \text{if } \max_{i \in \Omega_M} \rho_i < \tilde{\rho} \\ \beta & \text{otherwise} \end{cases}. \tag{47}$$

Under Assumptions 3 and 5, and for $L > 1$ and $\boldsymbol{\rho} \in \mathbb{P}_L(\Omega_M)$,

$$V_{\pi_{EJS}}(\boldsymbol{\rho}) \leq \frac{H(\boldsymbol{\rho}) + \max\{\log\log M, \log\log L\}}{\alpha} \tag{48}$$

$$+ \frac{\log L}{\beta} + \frac{3 \times 2^{2\xi_M + 4}}{\alpha\beta} + 1. \tag{49}$$

Note that the above theorems result in non-trivial upper bounds if one can find strictly positive values for α and β. In other words, the above theorems should be interpreted to lay out a proof methodology. In fact, in Section V, we use these results to prove the asymptotic optimality of a low-complexity search algorithm to find an object in a noisy environment. Next we provide an upper bound for $V_{\pi_{EJS}}$ for the general active hypothesis testing problem by characterizing α and β in (47) and using Corollary 2. For notational simplicity, let

$$I_0(M) := \max_{\boldsymbol{\lambda} \in \mathbb{P}(\mathcal{A}_M)} \min_{i \in \Omega_M} \min_{\hat{\boldsymbol{\rho}} \in \mathbb{P}_L(\Omega_M)} \sum_{a \in \mathcal{A}_M} \lambda_a D(q_i^a \| \sum_{j \neq i} \frac{\hat{\rho}_j}{1 - \hat{\rho}_i} q_j^a), \tag{50}$$

$$D_i(M) := \max_{\boldsymbol{\lambda} \in \mathbb{P}(\mathcal{A}_M)} \min_{\hat{\boldsymbol{\rho}} \in \mathbb{P}_L(\Omega_M)} \sum_{a \in \mathcal{A}_M} \lambda_a D(q_i^a \| \sum_{j \neq i} \frac{\hat{\rho}_j}{1 - \hat{\rho}_i} q_j^a), \ \forall i \in \Omega_M. \tag{51}$$

Note that for any $\boldsymbol{\lambda} \in \mathbb{P}(\mathcal{A}_M)$ and $\boldsymbol{\rho} \in \mathbb{P}_L(\Omega_M)$,

$$\max_{a \in \mathcal{A}_M} EJS(\boldsymbol{\rho}, a) \geq \sum_{a \in \mathcal{A}_M} \lambda_a EJS(\boldsymbol{\rho}, a)$$

$$\geq \min_{\hat{\boldsymbol{\rho}} \in \mathbb{P}_L(\Omega_M)} \sum_{a \in \mathcal{A}_M} \lambda_a EJS(\hat{\boldsymbol{\rho}}, a), \tag{52}$$

which implies that

$$\max_{a \in \mathcal{A}_M} EJS(\boldsymbol{\rho}, a) \geq I_0(M). \tag{53}$$

Similarly, we can show that for any $\boldsymbol{\lambda} \in \mathbb{P}(\mathcal{A}_M)$ and $\boldsymbol{\rho} \in \mathbb{P}_L(\Omega_M)$ such that $\rho_i \geq \tilde{\rho}$,

$$\max_{a \in \mathcal{A}_M} EJS(\boldsymbol{\rho}, a)$$

$$\geq \sum_{a \in \mathcal{A}_M} \lambda_a EJS(\boldsymbol{\rho}, a)$$

$$\geq \min_{\hat{\boldsymbol{\rho}} \in \mathbb{P}_L(\Omega_M)} \sum_{a \in \mathcal{A}_M} \lambda_a \tilde{\rho} D(q_i^a \| \sum_{j \neq i} \frac{\hat{\rho}_j}{1 - \hat{\rho}_i} q_j^a), \tag{54}$$

which implies that

$$\max_{a \in \mathcal{A}_M} EJS(\boldsymbol{\rho}, a) \geq \tilde{\rho} D_i(M) \quad \text{if } \rho_i \geq \tilde{\rho}. \tag{55}$$

Corollary 2 together with (53) and (55) provides the following upper bound on the expected total cost of policy π_{EJS},

$$V_{\pi_{EJS}}(\boldsymbol{\rho}) \leq \frac{H(\boldsymbol{\rho}) + \max\{\log\log M, \log\log L\}}{I_0(M)} + \frac{\log L}{\tilde{\rho} \min_{i \in \Omega_M} D_i(M)} + \frac{3 \times 2^{2\xi_M + 4}}{\tilde{\rho} \min_{i \in \Omega_M} D_i(M) I_0(M)} + 1. \tag{56}$$

Next we consider policy π_2, a two-phase policy which was introduced and analyzed in [15]. Let $\boldsymbol{\eta}_0$ and $\boldsymbol{\eta}_i$, $i \in \Omega_M$, be vectors in $\mathbb{P}(\mathcal{A}_M)$ that achieve the maximum in (50) and (51), respectively, i.e.,

$$\boldsymbol{\eta}_0 := \arg\max_{\boldsymbol{\lambda} \in \mathbb{P}(\mathcal{A}_M)} \min_{i \in \Omega_M} \min_{\hat{\boldsymbol{\rho}} \in \mathbb{P}_L(\Omega_M)} \sum_{a \in \mathcal{A}_M} \lambda_a D(q_i^a \| \sum_{j \neq i} \frac{\hat{\rho}_j}{1 - \hat{\rho}_i} q_j^a), \tag{57}$$

$$\boldsymbol{\eta}_i := \arg\max_{\boldsymbol{\lambda} \in \mathbb{P}(\mathcal{A}_M)} \min_{\hat{\boldsymbol{\rho}} \in \mathbb{P}_L(\Omega_M)} \sum_{a \in \mathcal{A}_M} \lambda_a D(q_i^a \| \sum_{j \neq i} \frac{\hat{\rho}_j}{1 - \hat{\rho}_i} q_j^a). \tag{58}$$

Moreover, let η_{0a} and η_{ia} denote elements of $\boldsymbol{\eta}_0$ and $\boldsymbol{\eta}_i$ corresponding to $a \in \mathcal{A}_M$, respectively. Consider a threshold

$\check{\rho}$, $\check{\rho} > \frac{1}{2}$. Markov (randomized) policy π_2 is defined as follows:

- If $\rho_i \geq 1 - L^{-1}$, retire and select H_i as the true hypothesis;
- If $\rho_i \in [\check{\rho}, 1 - L^{-1})$, then $\pi_2(a|\boldsymbol{\rho}) = \eta_{ia}$, $\forall a \in \mathcal{A}_M$;
- If $\rho_i < \min\{\check{\rho}, 1 - L^{-1}\}$, for all $i \in \Omega_M$, then $\pi_2(a|\boldsymbol{\rho}) = \eta_{0a}$, $\forall a \in \mathcal{A}_M$.

Next theorem provides an upper bound on the expected total cost of π_2:

Theorem 7 (see [15])**.** *Under Assumptions 3 and 5, and for $L > 1$ and any $\boldsymbol{\rho} \in \mathbb{P}_L(\Omega_M)$,*

$$V_{\pi_2}(\boldsymbol{\rho}) \leq \frac{H(\boldsymbol{\rho}) + \xi_M + K_2''}{I_0(M)} + \sum_{i=1}^{M} \rho_i \frac{\log L}{D_i(M)} + 1,$$

where K_2'' is a constant independent of L and M.

The proof of Theorem 7 relies on the (non-Bayesian) analysis of the conditional expected cost. Following a similar approach and for large values of L and M, the upper bound (56) for π_{EJS} can be tightened as follows:

Proposition 2. *Under Assumptions 3 and 5, and for $L > 1$ and $\boldsymbol{\rho} \in \mathbb{P}_L(\Omega_M)$,*

$$V_{\pi_{EJS}}(\boldsymbol{\rho}) \leq \frac{H(\boldsymbol{\rho}) + \max\{\log\log M, \log\log L\}}{I_0(M)} + \sum_{i=1}^{M} \rho_i \frac{\log L + \frac{K''_{EJS} \times 2^{2\xi_M}}{I_0(M)}}{D_i(M)} + 1,$$

where K''_{EJS} is a constant independent of L and M.

Note that an upper bound for the class of sequential adaptive policies can be provided by analyzing the performance of a heuristic two-phase policy π_{SA} which can be described as follows:

- If $\rho_i \geq 1 - L^{-1}$, retire and select H_i as the true hypothesis;
- If $\rho_i \in [\check{\rho}, 1 - L^{-1})$, then $\pi_{SA}(a|\boldsymbol{\rho}) = \mu_{ia}$, $\forall a \in \mathcal{A}_M$;
- If $\rho_i < \min\{\check{\rho}, 1 - L^{-1}\}$, for all $i \in \Omega_M$, then $\pi_{SA}(a|\boldsymbol{\rho}) = \mu_{0a}$, $\forall a \in \mathcal{A}_M$;

where $\boldsymbol{\mu}_0$ and $\boldsymbol{\mu}_i$, $i \in \Omega_M$, are vectors in $\mathbb{P}(\mathcal{A}_M)$ that achieve the maximum in (59) and (60), respectively:

$$\tilde{I}_0(M) := \max_{\boldsymbol{\lambda} \in \mathbb{P}(\mathcal{A}_M)} \min_{i \in \Omega_M} \min_{j \neq i} \sum_{a \in \mathcal{A}_M} \lambda_a D(q_i^a \| q_j^a), \quad (59)$$

$$\tilde{D}_i(M) := \max_{\boldsymbol{\lambda} \in \mathbb{P}(\mathcal{A}_M)} \min_{j \neq i} \sum_{a \in \mathcal{A}_M} \lambda_a D(q_i^a \| q_j^a), \ \forall i \in \Omega_M. \quad (60)$$

The next proposition provides a benchmark for comparison with π_{EJS}.

Proposition 3. *Under Assumptions 3 and 5, and for $L > 1$, $\boldsymbol{\rho} \in \mathbb{P}_L(\Omega_M)$, and arbitrary $\iota \in (0, 1)$,*

$$V_{\pi_{SA}}(\boldsymbol{\rho}) \leq \frac{H(\boldsymbol{\rho}) + \log M + K''}{\tilde{I}_0(M)}(1 + \iota) + \sum_{i=1}^{M} \rho_i \frac{\log L}{\tilde{D}_i(M)}(1 + \iota) + \left(6M + \frac{M}{(\frac{\iota/2}{1+\iota})^5 (\frac{\tilde{I}_0(M)}{2\xi_M})^4}\right) \times \left(L(1 - \max_{j \in \Omega_M} \rho_j)\right)^{-\frac{\iota^3}{(1+\iota)^2} \frac{\tilde{I}_0^2(M)}{4\xi_M^3}},$$

where K'' is a constant independent of L and M.

E. Discussion: Contributions and Prior Work

The problem of active sequential hypothesis testing naturally arises in a broad spectrum of applications such as medical diagnosis [16], cognition [17], sensor management [18], underwater inspection [19], generalized search [20], group testing [21], and channel coding with perfect feedback [22].

The most well known instance of our problem is the case of binary hypothesis testing with passive sensing ($M = 2$, $K = 1$), first studied by Wald [23]. In this instance of the problem, the optimal action at any given time is provided by a sequential probability ratio test (SPRT). There are numerous studies on the generalizations to $M > 2$ ($K = 1$) and the performance of known simple and practical heuristic tests such as MSPRT [24], [25], [26].

Active hypothesis testing also generalizes another classic problem in the literature: the comparison of experiments first introduced by Blackwell [9]. This is a single-shot version of the active hypothesis testing problem in which the decision maker can choose one of several (usually two) actions/experiments to collect a single observation sample before making the final decision. There have been extensive studies [9], [27], [28], [10], [29], [30], [31] on comparing the actions. As we saw in previous sections, applying various results from [9], [10] in our context of active hypothesis testing and utilizing a dynamic programming interpretation, an optimal notion of information utility, i.e., an optimal measure to quantify the information gained by different sensing actions, can be derived [32]. Inspired by this view of the problem, which coincides with that promoted by DeGroot [33], we provided a set of (uniform) lower bounds for optimal information utility. Furthermore, two heuristic policies were developed.

The first attempt to solve **Problem-AHT** is Chernoff's work on active binary composite hypothesis testing [34] where a heuristic randomized policy was proposed and whose asymptotic performance was analyzed. More specifically, under a certain technical assumption on uniformly distinguishable hypotheses, the proposed heuristic policy is shown to achieve asymptotic optimality where the notion of asymptotic optimality [34] denotes the relative tightness of the performance upper bound associated with the proposed policy and the lower bound associated with the optimal policy. Given the similar setting of the problem, it is important for us to delve deeper into Chernoff's work and method of analysis. To achieve asymptotic optimality, Chernoff proposed the following randomized scheme to select actions: At each time t, find the most likely true hypothesis, and then select an action that can discriminate this hypothesis the best from each and every element in the set of alternative hypothesis. Much of the subsequent literature extended this approach [35], [36], [37], [38], [39], [40], [41]. Chernoff showed that as L goes to infinity, the relative difference between the expected total cost achieved by his proposed scheme and the optimal expected total cost approaches zero; which he termed as *asymptotic optimality*.[5] One of the main drawbacks of Chernoff's asymptotic optimality notion was his neglecting the complementary role of asymptotic analysis in M. In particular, the notion of asymptotic optimality in L falls short in showing the tension between using (asymptotically) large number of samples to discriminate among a few hypotheses with (asymptotically) high accuracy or a (asymptotically) large number of hypotheses with a lower degree of accuracy. As a result, although the scheme proposed in [34] and its subsequent extensions [35], [36], [37], [38], [39], [40], [41] are asymptotically optimal in L, their provable information acquisition rate is restricted to zero. Intuitively, the rate of information acquisition under any given heuristic relates to the ratio between $\log M$ and the expected number of samples: the larger this ratio the faster information is acquired.

To obtain asymptotic characterization of the optimal expected total cost in a non-zero rate regime, it is important to propose schemes which scale optimally with M as well. In his seminal paper [22], Burnashev tackled the primal (constrained) version of Problem (P) in the context of channel coding with feedback, and provided lower and upper bounds on the expected number of samples (or equivalently channel uses) required to convey one of M uniformly distributed messages over a discrete memoryless channel (DMC) with a desired probability of error. The lower bound identified the dominating terms in both number of messages and error probability, hence characterized the optimal reliability function (also known as the error exponent) in addition to the feedback capacity (which was known to coincide with the Shannon capacity [42]). In this paper, we generalize[6] this lower bound to the problem of active sequential hypothesis testing. In addition to a lower bound on an expected number of samples, Burnashev proposed a coding scheme with two phases of operation whose performance provides a tight upper bound (in both number of messages and error probability). It is interesting to note that the scheme of Chernoff, if specialized to channel coding with feedback, coincides with the confirmation (second) phase of Burnashev's scheme and is of a repetition code nature. This means that while the first phase of Burnashev's scheme is designed to achieve any information rate up to the (Shannon) capacity of the channel, Chernoff's one-phase scheme has a rate of information acquisition equal to zero. Our upper bounds discussed earlier are all nspired by Burnashev's coding scheme,

The above results have all been obtained under an important technical assumption: Assumption 2. This assumption is fairly limiting as it excludes the cases where the observation kernel has unbounded support. However, in our prior work, we have showed that this assumption can be significantly weakened [15].

V. Searching with Measurement-Dependent Noise

In this section, we consider a target search problem on a unit internal where at any given time an agent can choose a region to probe into for the presence of the target in that region. The measurements are assumed to be a noisy indicator of the presence of the target in the search region the agent probes. Searching with measurement independent noise is known [44] to be equivalent to the problem of channel coding over a Binary Symmetric Channel (BSC) with crossover probability p, where the adaptivity of the search strategy corresponds to the availability of noiseless feedback and the targeting rate corresponds to the channel capacity. It is well known that feedback cannot increase channel capacity [45] but can increase the reliability of data transmission [46], [47] with allowing variable time of termination. Correspondingly [48] demonstrates how to design an adaptive search algorithm that achieves the best possible targeting rate and reliability, with the help of the Extrinsic Jensen-Shannon (EJS) divergence [49], where searching rate is defined to be the ratio of the logarithm of the search resolution over the expected number of samples and the reliability is defined to be the ratio of the logarithm of error probability over the expected number of samples.

In many practical applications, however, the measurement noise depends on the size of the region probed. In particular, the search of a larger region is prone to more noise than probing smaller regions. In other words, the crossover probability (false alarm and miss detection) associated with searching region of size α can be thought of a non-decreasing function $p[\alpha]$. In this setting, adaptivity is known to be essential [3] and can increase not only reliability but also the targeting rate compared with non-adaptive strategies. More specifically, the analysis in [3] quantifies the adaptivity gain in realistic setting when the noise $p[\alpha]$ is decreasing with the search size α. The achievable adaptive search algorithm in [3], however, is a rudimentary three-phase search scheme constructed primarily for analytical/proof purposes. In particular, the first two phases of the achievable scheme are non-adaptive and based on a random partitioning of the search area, and the third phase consists of an ACK/NACK operation. While this simple hueristics allows for a simple proof of adaptivity gain and is shown to achieve the asymptotically optimal performance, we can achieve this optimal performance via a truly adaptive and deterministic policy

[5] In [34], the objective was to minimize $c\mathbb{E}[\tau]+\text{Pe}$ and the proposed policy was shown to be asymptotically optimal as $c \to 0$. It is straightforward to show that for $L = \frac{1}{c}$, this problem coincides with Problem (P) defined in this paper. However, we have chosen $\mathbb{E}[\tau] + L\text{Pe}$ as an objective function for Problem (P) because of its interpretation as the Lagrangian relaxation of an information acquisition problem in which the objective is to minimize $\mathbb{E}[\tau]$ subject to $\text{Pe} \leq \epsilon$ where $\epsilon > 0$ denotes the desired probability of error.

[6] In [43], Burnashev attempted to tackle the problem of active sequential hypothesis testing by Chernoff [34]. However, the sensing actions in [43] were allowed to be functions of the true hypothesis, θ, which, in general, is not observable in the active testing setting [34]. In this sense, [43] only extends Burnashev's earlier work [22] on variable-length coding over a discrete memoryless channel (DMC) with feedback to allow for more general channels.

with non-asymptoticaly good performance. Furthermore, the implementation of three-phase search scheme of [3] requires tuning several parameters which impact the performance especially in the non-asymptotic regime.

In this section, we proposed a deterministic single-phase search strategy that is low complexity and is proven to achieve the best possible targeting rate and reliability under measurement dependent noise. Practically, unlike the three-phase scheme in [3], our proposed algorithm is universal and can be easily implemented at any given resolution. Furthermore, we show, numerically, that the proposed algorithm has a superior non-asymptotic performance.

A. Problem Setup

The problem setup is similar to that of [3]. Let $\theta \in [0,1)$ be the fixed position of the target, uniformly placed on the unit interval. We wish to estimate the target position θ. At time t, an agent may seek the target by choosing (possibly at random) any measurable *query set* $Q_t \subset [0,1)$ to probe. Without loss of generality, we will assume throughout that $|Q_t| \leq \frac{1}{2}$ almost surely. Let $X_t = \mathbb{1}(\theta \in Q_t)$ denote the clean binary signal indicating whether the target is in the probed region. The agent obtains a corrupted version Y_n of X_n, with noise level that corresponds to the size of the region Q_t. Specifically,

$$Y_t = X_t \oplus Z_t, \tag{61}$$

where $Z_t \sim \text{Bern}\left(p[|Q_t|]\right)$ and $p : (0, 1/2] \mapsto [0, 1/2)$ is a continuous and non-decreasing function.

A *search strategy* is a causal protocol for determining the sequence of search region, Q_t, $t = 1, 2, 3, ...$, the stopping time τ, and estimators $\hat{\theta}_\tau$ of the position. A strategy is said to be *non-adaptive* if the choice of the region Q_t is independent of Y^{t-1} and the stopping time is fixed in advance. Otherwise, the strategy is said to be *adaptive*. A strategy is said to have *search resolution* δ and error probability ε if for any θ,

$$\mathbb{P}(\ |\ \hat{\theta}_\tau - \theta\ |\ \leq \delta) \geq 1 - \varepsilon. \tag{62}$$

Let $\tau_{\epsilon,\delta}$ denote the stopping time that a given search strategy declares the target location with resolution δ, error probability ϵ. We are interested in characterizing the expected search time $\tau_{\epsilon,\delta}$.

B. Finite Resolution Search Strategies

Let us divide the unit interval $[0, 1]$ into $1/\delta$ sub-intervals (referred to as search bins), where $1/\delta$ is an integer. We consider search strategies that choose the search subset Q_t to coincide with a subset of the search bins. And instead of declaring a location $\hat{\theta}$, we declare only the index of the bin $\{1, 2, ..., \frac{1}{\delta}\}$ which is believed to hold the target. Let i^θ denote the bin index that contains the true position of the target. It is clear that if a finite resolution δ search strategy satisfies $\mathbb{P}(\hat{i} = i^\theta) \geq 1-\varepsilon$, then it achieves targeting resolution δ with reliability ϵ.

Furthermore, it is sufficient for us to consider strategies that select the next search region $S_t \subseteq \{1, 2, ..., \frac{1}{\delta}\}$ based on the following posterior probability

$$\rho_i(t) = \mathbb{P}(i_\theta = i \mid \mathcal{F}_t), \quad i = 1, 2, ..., 1/\delta. \tag{63}$$

C. Prior work

The problem of noisy search with measurement independent noise was first introduced in [48], where non-asymptotic lower bound and (loose) upper bound on $\mathbb{E}[\tau_{\epsilon,\delta}]$ was derived, i.e.

$$\begin{aligned}\frac{\log(1/\delta)}{C(0)}+\frac{\log(1/\epsilon)}{C_1(\delta)}+o\left(\log\frac{1}{\delta\epsilon}\right) &\leq \mathbb{E}[\tau_{\epsilon,\delta}] \\ &\leq \ \frac{\log(1/\delta)}{C(1/2)}+\frac{\log(1/\epsilon)}{C_1(\delta)}+o\left(\log\frac{1}{\delta\epsilon}\right).\end{aligned} \tag{64}$$

In [3], an optimal[7] family of three-phase adaptive scheme was shown, for all $\alpha > 0$, to achieve

$$\mathbb{E}[\tau_{\epsilon,\delta}] \leq \ \frac{\log(1/\delta)}{C(\alpha)}+\frac{\log(1/\epsilon)}{C_1(\delta)}+o\left(\log\frac{1}{\delta\epsilon}\right), \tag{65}$$

for $\delta \ll \alpha$. In particular, each scheme in this class consists of a first phase, where random collection of bins are searched over time until the smallest number of bins that hold $1-\epsilon/2$ posterior probability is no more than α/δ. Phase 2 and 3 of the strategy is then searching over these α/δ bins to ensure that the observation noise is below $p[\alpha]$. Assuming a worse case scenario where the observation noise in phase 2 and phase 3 is Bernoulli with flip probability exactly equal to $p[\alpha]$ (instead of upper-bounded by it), the problem can be mapped to the problem of channel coding with feedback, where the two-phase Yamamoto-Itoh [50] scheme is applied (to the α/δ bins with the target posterior identified in the first phase). While this three-phase scheme achieves the theoretical limits of performance, the design of a low complexity, deterministic, sequential and adaptive searching scheme remains.

D. Sorted Posterior Matching: overview of the results

In contrast to the three-phase scheme in [3], here we propose a low complexity single-phase scheme that is easy to implement and is proven to achieve targeting rate and reliability arbitrarily close to $C(0)$ and $C_1(0)$. The single-phase scheme we propose here, denoted by $sortPM$, is a modification of the Posterior Matching algorithm [51]. In particular, we apply the procedure of the deterministic Generalized Horstein-Burnashev-Zigangirov Scheme in [49] on $\boldsymbol{\rho}^{\downarrow}(t)$, selecting

$$S_t^{sortPM} = \{i : \rho_i(t) \geq \rho^{\downarrow}_{k^*}(t)\}, \tag{66}$$

where

$$k^* = \arg\min_k \left|\sum_{i=1}^{k} \rho_i^{\downarrow}(t) - \frac{1}{2}\right|. \tag{67}$$

[7]when the noise function $p[\cdot]$ is strictly increasing, the scheme is significantly better than non-adaptive rate maximization scheme.

Remark 6. Note that the complexity of $sortPM$ search is $O(\frac{1}{\delta}\log\frac{1}{\delta})$ in each step.

Theorem 8. *For any $\alpha \gg \delta > 0$, the expected search time of $sortPM$ is bounded by*

$$\mathbb{E}[\tau_{\epsilon,\delta}] \leq \frac{\log(1/\delta)}{C(\alpha)} + \frac{\log(1/\epsilon)}{C_1(\delta)} + o\left(\log\frac{1}{\delta\epsilon}\right), \quad (68)$$

as $(\delta,\epsilon) \to 0$

This result shows that we can achieve the best possible targeting rate and error exponent using a simple single-phase scheme without the implementation difficulty associated with choosing the transition parameter α for the three-phase phase adaptive scheme in [3].

E. Analysis: Upper Bound

The posterior $\boldsymbol{\rho}(t),\ t = 0, 1, 2, ...,$ forms a Markov chain since the search set S_t only depends on $\boldsymbol{\rho}(t)$. However, it is difficult to analyze the transition probability and the behaviour of this Markov Chain directly. Similar to [49], we consider the functional average log-likelihood and exploit the Extrinsic Jensen-Shannon divergence (EJS) [49] to analyze our $sortPM$ search algorithm and then give a complete proof for Theorem 8.

In this subsection, we use Theorem 6 and Corellary 2 for analyzing the random drift of average likelihood from time 0 with the initial value $U(0)$ up to the first crossing time $\nu := \min\{t : U(t) \geq \log\frac{1}{\epsilon}\}$ which is closely related to the expected drift given by EJS. In particular, we can then establish an upper bound for the expected target time $\mathbb{E}[\tau_{\epsilon,\delta}]$ in terms of the predefined error probability ϵ and the resolution δ as well as the guaranteed lower bound on instantenouseEJS. Specifically we have the following theorem:

Fact 5 (Theorem 1 in [49])**.** *Define*

$$\tilde{\rho} := 1 - \frac{1}{1 + \max\{\log(1/\delta), \log(1/\epsilon)\}}. \quad (69)$$

For adaptive search strategy with search region S_t, if

$$EJS(\boldsymbol{\rho}(t), S_t) \geq R \quad \forall t \geq 0 \quad (70)$$

and

$$EJS(\boldsymbol{\rho}(t), S_t) \geq \tilde{\rho}E \quad \forall t \geq 0 \text{ s.t. } \max_i \rho_i(t) \geq \tilde{\rho}, \quad (71)$$

we have the expected stopping time associated with error probability ϵ and resolution δ bounded by

$$E[\tau_{\epsilon,\delta}] \leq \frac{\log(1/\delta)}{R} + \frac{\log(1/\epsilon)}{E} + o\left(\log\frac{1}{\delta\epsilon}\right) \quad (72)$$

as $(\delta,\epsilon) \to 0$.

The proof follows similarly with that of [Theorem 1, [49]]. Furthermore, along with the proof of [Proposition 3, [49], we can show that $EJS(\boldsymbol{\rho}(t), S_t^{sortPM}) \geq C(|S_t|)$. Straightforwardly, we have $EJS(\boldsymbol{\rho}(t), S_t^{sortPM}) \geq C(1/2)$ for all t and $EJS(\boldsymbol{\rho}(t), S_t^{sortPM}) \geq C_1(\delta)$ if $\max_i \rho_i(t) \geq \tilde{\rho}$. To prove Theorem 8, however, we will need to compare $EJS(\boldsymbol{\rho}(t), S_t^{sortPM})$ with $C(\alpha)$ for arbitrarily small α for all t. In other words, the difficulty of proving achievability lies in the fact that the expected drift $EJS(\boldsymbol{\rho}(t), S_t^{sortPM})$ is only larger than $C(|S_t|)$ but not $C(\alpha)$ for arbitrary α for all time. Our approach here is to show that by the operation of $sortPM$, the search size $|S_t|$ will shrink in and spend the majority of time within a region of size α (Lemma 4). To prove this, we binned the sorted posterior into multiple bins of size α and consider another averaged log-likehood function $U_\alpha(t)$ defined upon the size-α bins. The event $\{|S_t| \leq \alpha\}$ is then guaranteed by $\{U_\alpha(t) \geq 0\}$ whose probability is established via Azuma's inequality on the submartingale $U_\alpha(t)$ and the lower bound of its expected drift (a modified EJS) provided. In the next section, we provide further the outline of proof.

F. Sketch of the Proof of Theorem 8

To prove that the search strategy $sortPM$ can achieve $C(\alpha)$ for arbitrary small α, we consider the event that the S_t^{sortPM} picks up the corresponding search size $|Q_t^{sortPM}| \leq \alpha$. This happens if

$$\rho_1^\alpha(t) := \sum_{i=1}^{\lfloor \alpha/\delta \rfloor} \rho_i^{\downarrow}(t) \geq \frac{1}{2}. \quad (73)$$

Define $E_t = \{\rho_1^\alpha(t) < \frac{1}{2}\}$, and $F_n = \bigcup_{t=n}^{\infty} E_t$. Under event F_n^C, we guarantee that after time n the search size of $sortPM$ is always smaller than α and hence the probability of crossover noise is limited by $p[\alpha]$. In other words,

$$\begin{aligned}
\mathbb{E}[\tau_{\epsilon,\delta}] &= \int_\Omega \tau_{\epsilon,\delta}\ d\mathbb{P} \leq \sum_{t=n}^{\infty} \int_{E_t} \tau_{\epsilon,\delta}\ d\mathbb{P} + \int_{F_n^C} \tau_{\epsilon,\delta}\ d\mathbb{P} \\
&= \sum_{t=n}^{\infty} \int_{E_t} \mathbb{E}[\tau_{\epsilon,\delta} \mid \boldsymbol{\rho}(t)]\ d\mathbb{P} + \int_{F_n^C} \tau_{\epsilon,\delta}\ d\mathbb{P} \\
&\overset{(a)}{\leq} \sum_{t=n}^{\infty} \mathbb{P}(E_t)\left(t + \frac{\log\frac{1}{\delta}}{C(0.5)} + \frac{\log\frac{1}{\epsilon}}{C_1(\delta)} + o\left(\log\frac{1}{\delta\epsilon}\right)\right) \\
&\quad + \int_{F_n^C} \tau_{\epsilon,\delta}\ d\mathbb{P} \\
&\overset{(b)}{\leq} \sum_{t=n}^{\infty} \mathbb{P}(E_t)\left(t + \frac{\log\frac{1}{\delta}}{C(0.5)} + \frac{\log\frac{1}{\epsilon}}{C_1(\delta)} + o\left(\log\frac{1}{\delta\epsilon}\right)\right) \\
&\quad + n + \frac{\log\frac{1}{\delta}}{C(\alpha)} + \frac{\log\frac{1}{\epsilon}}{C_1(\delta)} + o\left(\log\frac{1}{\delta\epsilon}\right).
\end{aligned} \quad (74)$$

Here (a) follows from the time homogeneity of the Markov Chain $\boldsymbol{\rho}(t)$ together with Fact 5, written as

$$\mathbb{E}[\tau_{\epsilon,\delta} \mid \boldsymbol{\rho}(t)] \leq t + \frac{\log\frac{1}{\delta}}{C(0.5)} + \frac{\log\frac{1}{\epsilon}}{C_1(\delta)} + o\left(\log\frac{1}{\delta\epsilon}\right). \quad (75)$$

Similar arguments can be made for (b),

$$\int_{F_n^C} \tau_{\epsilon,\delta}\ d\mathbb{P} \leq n + \frac{\log\frac{1}{\delta}}{C(\alpha)} + \frac{\log\frac{1}{\epsilon}}{C_1(\delta)} + o\left(\log\frac{1}{\delta\epsilon}\right). \quad (76)$$

Now the problem reduces to finding an appropriate upper-bound for $\mathbb{P}(E_t)$, provided in the lemma below.

Lemma 4. *Using the search strategy $sortPM$, we have*

$$\mathbb{P}(E_t) := \mathbb{P}\left(\rho_1^{\alpha}(t) < \frac{1}{2}\right) < k_f e^{-tE_0} \quad \forall t > T_0 \tag{77}$$

for some $T_0, E_0 > 0$.

This lemma together with (74) implies that

$$\mathbb{E}[\tau_{\epsilon,\delta}] \leq O(n) + \frac{k_f e^{-nE_0}}{1-e^{-E_0}}\left(\frac{\log\frac{1}{\delta}}{C(0.5)} + \frac{\log\frac{1}{\epsilon}}{C_1(\delta)}\right) + \frac{\log\frac{1}{\delta}}{C(\alpha)} + \frac{\log\frac{1}{\epsilon}}{C_1(\delta)} + o\left(\log\frac{1}{\delta\epsilon}\right), \quad n > T_0. \tag{78}$$

Finally letting $n = \lceil \log\log\frac{1}{\delta\epsilon}\rceil$ in equation (78), we have the assertion of the theorem.

VI. Bayesian Active Learning with non-Persistent Noise

In this section, we consider the problem of noisy Bayesian active learning where we are given a finite set of functions $\mathcal{H}$, a sample space $\mathcal{X}$, and a label set $\mathcal{L}$. One of the functions in $\mathcal{H}$ assigns labels to samples in $\mathcal{X}$, and our goal is to identify this function when the result of a label query on a sample is corrupted by independent noise. The objective is to declare one of the functions in $\mathcal{H}$ as the true label generating function with high confidence using as few label queries as possible, by selecting the queries adaptively and in a strategic manner.

A special case of the problem, first considered by [52], arises when the label set is binary and the natural sampling strategy for Bayesian active learning becomes closely related to Generalized Binary Search (GBS). In the binary label setting, GBS queries the label of a sample x for which the size of the subsets of functions that label x as $+1$ and -1 respectively, are as balanced as possible. A variant of GBS is Modified Soft-Decision Generalized Binary Search (MSGBS), which was introduced by [52] to address the case when the observed labels may be noisy. [52] analyzes the performance of MSGBS, under a symmetric and non-persistent noise model which flips the labels randomly, and shows that the (fixed) number of samples required to identify the correct function with probability of error satisfying Pe $\leq \epsilon$ is $\frac{\log M + \log\frac{1}{\epsilon}}{\lambda}$, where M is the number of functions in the class $\mathcal{H}$, and λ is a parameter which depends on the structure of the function class, the sample space, and the noise rate. The mainfirst contribution of this paper is to generalize the above problem to the case of general (non-binary) label set with general (and potentially non-symmetric) non-persistent observation noise.

By allowing for the number of samples collected to be determined in a sequential manner (according to a random stopping time as a function of past observations), we draw a parallel between active sequential hypothesis testing [15] and Bayesian active learning. In active sequential hypothesis testing, we are given a set of M hypotheses, and a set of actions; each action, conditioned on the true hypothesis, has a certain probability of yielding an outcome. We observe that Bayesian active learning is a special case of active hypothesis testing, where the hypotheses map to functions, actions map to samples, and the outcomes map to noisy observation of labels. This view of the problem allows for a natural extension of the model of [52] to the non-binary Bayesian active learning setting, where the label noise might be label dependent and asymmetric. Relying on this connection, we derive a universal lower bound on the *expected* number of samples required to identify the true hypothesis among M with reliability ϵ as a function of noise model parameters. Furthermore, we can take advantage of the connection between Bayesian learning and active sequential hypothesis testing to proposed an active sequential hypothesis test, MaxEJS, that, at each step, selects the action that maximizes the EJS divergence. In special cases where $\mathcal{H}$ is locally identifiable and sample rich, we provide upper bounds on the query complexity of MaxEJS. These bounds establish the asymptotic optimality of MaxEJS for sample rich function class. Also, similar optimality bounds are shown when $\mathcal{H}$ is specialized further to the class of threshold binary valued functions.

A. Problem Setup

In this section, we provide the mathematical description of the problem of Bayesian active learning.

> **Problem-NBAL** [Noisy Bayesian Active Learning]
> In the Bayesian active learning problem, we are given a *sample space* $\mathcal{X}$, a finite label set $\mathcal{L}$, and a finite *observation space* $\mathcal{Y}$. We are also given a set $\mathcal{H} = \{h_1, h_2, \ldots, h_M\}$ of M distinct functions, where each $h_i : \mathcal{X} \to \mathcal{L}$ maps elements in the sample space $\mathcal{X}$ to the label set $\mathcal{L}$. We assume that one of the functions in $\mathcal{H}$, denoted by h_θ, produces the correct labeling on $\mathcal{X}$. The decision maker is allowed to *query* samples from $\mathcal{X}$. Querying a sample x generates an observation in $y \in \mathcal{Y}$ whose distribution is a given function of the true label as determined by the function h_θ. More specifically, if h_θ is the true underlying function and hence $l = h_\theta(x)$ is the true label of sample x, then the result of a query on x is a $\mathcal{Y}$-valued random variable with probability mass function $f_l(\cdot)$. We assume that $\{f_l(\cdot)\}_{l\in\mathcal{L}}$ are fixed and known, and observations are conditionally independent over time.[8]
> The goal of the decision maker is to determine the identity of the function in $\mathcal{H}$ that generates the true labels by an adaptive sequential query of a small number of samples. We assume that the decision maker does not have any extra prior knowledge on the identity of the true function; in other words, it begins with a uniform prior over $\mathcal{H}$. Let τ be the stopping time at which the decision maker retires and declares the label generating

[8]For notational convenience, we present our results in this paper for a finite observation space $\mathcal{Y}$. However, the results could be easily extended to general $\mathcal{Y}$ where $\{f_l(\cdot)\}_{l\in\mathcal{L}}$ are density functions with respect to some σ-finite measure ν.

function $h_{\hat{\theta}}$. Furthermore, let $\text{Pe} = P(\hat{\theta} \neq \theta)$ where θ is the index of the true function. In Bayesian active learning, the objective is to design a strategy for the decision maker for querying samples in $\mathcal{X}$ such that, for any given $\epsilon > 0$, we have

$$\text{minimize } \mathbb{E}[\tau] \text{ subject to } \text{Pe} \leq \epsilon. \tag{79}$$

Here the minimization is taken over the choice of the stopping time τ and the learning strategy and the expectation is taken with respect to the observation distribution as well as the Bayesian uniform prior on the true function in $\mathcal{H}$.

Let the decision maker's posterior belief about each possible function index $i \in \Omega$, updated after each sample query and observation for $t = 0, 1, \ldots, \tau$, be

$$\rho_i(t) := P(\{\theta = i\} | X^{t-1}, Y^{t-1}), \tag{80}$$

where

$$X^{-1} := \emptyset,\ X^{t-1} := [X(0), X(1), \ldots, X(t-1)] \quad \forall t \geq 1,$$
$$Y^{-1} := \emptyset,\ Y^{t-1} := [Y(0), Y(1), \ldots, Y(t-1)] \quad \forall t \geq 1.$$

We note that the dynamics of the information state, i.e., the posterior, follows Bayes' rule. But before we make this more precise, let us consider an alternative representation of querying a sample $x \in \mathcal{X}$:

Definition 3. A sample $x \in \mathcal{X}$ generates a $|\mathcal{L}|$-partition $\Xi^x := \{H_l^x\}_{l \in \mathcal{L}}$ of the function class, where $H_l^x = \{h \in \mathcal{H} : h(x) = l\}$, i.e., $\mathcal{H} = \cup_{l \in \mathcal{L}} H_l^x$.

This view allows us to characterize the observation distribution given the belief vector $\boldsymbol{\rho}$ and queried sample x as

$$f_x^{\boldsymbol{\rho}}(y) := \sum_{i \in \Omega} \rho_i f_{h_i(x)}(y) = \sum_{l \in \mathcal{L}} f_l(y) \sum_{i: h_i \in \mathcal{H}_l^x} \rho_i. \tag{81}$$

Therefore, given the belief vector $\boldsymbol{\rho}(t)$, querying sample $X(t) = x$ and observing (noisy) label $Y(t) = y$ results in a refinement of the posterior according to the Bayes' rule, i.e.,

$$\boldsymbol{\rho}(t+1) = \boldsymbol{\Phi}^x(\boldsymbol{\rho}(t), y) \tag{82}$$

where

$$\boldsymbol{\Phi}^x(\boldsymbol{\rho}, y) := \left[\rho_1 \frac{f_{h_1(x)}(y)}{f_x^{\boldsymbol{\rho}}(y)}, \ldots, \rho_M \frac{f_{h_M(x)}(y)}{f_x^{\boldsymbol{\rho}}(y)}\right]. \tag{83}$$

As a consequence of a connection between Bayesian active learning and the more general problem of information acquisition considered in the earlier chapters, we will analyze the problem via Extrinsic Jensen–Shannon divergence [53] which in this case can be written as:

$$EJS(\boldsymbol{\rho}, x) := \sum_{l \in \mathcal{L}} \sum_{i: h_i \in \mathcal{H}_l^x} \rho_i D\left(f_l \,\middle\|\, \frac{f_x^{\boldsymbol{\rho}} - \rho_i f_l}{1 - \rho_i}\right). \tag{84}$$

B. Analysis: Lower Bound

We now provide the first set of results – lower bounds on the optimal number of queries to identify the true function with high accuracy. Note that we expect the query complexity of our problem to depend on the characterizations of the discrete memoryless communication channel (DMC) which corrupts the true label's observations. This is a DMC with input alphabet set $\mathcal{L}$, output alphabet set $\mathcal{Y}$, and a collection of conditional probabilities $f_l(\cdot)$, $l \in \mathcal{L}$. We begin with a few assumptions on this channel.

Assumption 3. $C := \min_{g \in \mathbb{P}(\mathcal{Y})} \max_{l \in \mathcal{L}} D(f_l \| g) > 0.$

Assumption 4. $C_1 := \max_{k,l \in \mathcal{L}} D(f_k \| f_l) < \infty.$

Assumption 5. $C_2 := \max_{k,l \in \mathcal{L}} \sup_{y \in \mathcal{Y}} \frac{f_k(y)}{f_l(y)} < \infty.$

Note that C defined above is nothing but the Shannon capacity of the DMC with the collection of conditional probabilities $P(Y = y | L = l) = f_l(y)$, $l \in \mathcal{L}$ (See [54, Theorem 13.1.1]). In particular, the minimum is achieved by g^*, a convex combination of $\{f_l\}_{l \in \mathcal{L}}$, i.e., $g^* = \sum_{l \in \mathcal{L}} \pi_l^\star f_l$ where $\{\pi_l^\star\}_{l \in \mathcal{L}}$ is referred to as the *capacity-achieving input distribution* and has the property that for each $k \in \mathcal{L}$, if $\pi_k^\star > 0$, then $D(f_k \| g^*) = C$ (See [55, Theorem 4.5.1]). If Assumption 3 does not hold, that is if $C = 0$, the label queries will be completely noisy and no information can be retrieved from the label queries regarding the true function. In this sense, Assumption 3 is a necessary condition that is required for Problem (P) to have a meaningful solution.

Parameter C_1 emerges as an important quantity in the problem of variable-length coding with feedback: It denotes the maximum exponential decay rate of the error probability [22]. It is straight forward to show that $C \leq C_1$ and hence, Assumptions 3 and 4 imply that also $C_1 > 0$ and $C < \infty$.

Since, in general, $C_1 \leq \log C_2$, Assumption 4 is redundant with respect to Assumption 5. For observation distributions with finite support, i.e., when $|\mathcal{Y}| < \infty$, Assumption 5 ensures that the conditional distributions f_l, $l \in \mathcal{L}$, are absolutely continuous with respect to each other. Thus for observation distributions with finite support, Assumption 5 is a necessary and sufficient condition to ensure Assumption 4. On the other hand, for observation distributions with unbounded support, Assumption 5, which is stronger than Assumption 4, is a technical assumption made for ease of analysis, and will help us construct strong non-asymptotic bounds in closed form.

While the (non-asymptotic) bounds and analysis in this paper are all obtained under Assumptions 3 and 5, we have chosen to separately state Assumptions 4 and 5 in order to point out that it is possible to relax Assumption 5. More specifically, it is shown in [15] that at the cost of increasing notation, more complicated analysis, and loosening the non-asymptotic bounds, it is possible to relax Assumption 5 and obtain similar asymptotic characterizations only under Assumption 3 and a slightly stronger variant of Assumption 4.

The following lower bound on the minimum expected number of samples required to achieve $\mathrm{Pe} \leq \epsilon$ relies on dynamic programming methods similar to those used in the proof of Theorem 4. This lower bound is similar to [43, Theorem 1], [56, Theorem 1], and [57, Theorem 6].

Theorem 9. *Consider Problem (P) under Assumptions 3 and 5.*

$$\mathbb{E}[\tau_\epsilon^*] \geq \left[\frac{(1 - \frac{5}{\log \frac{4}{\epsilon}} - \frac{\epsilon}{2} \log \frac{4}{\epsilon}) \log M - \frac{8C_2}{\log \frac{4}{\epsilon}} - 4}{C} + \frac{\log \frac{1-\epsilon}{\epsilon} - 2 \log\log \frac{4}{\epsilon} - \log C_2 - 1}{C_1} \right]^+ . \tag{85}$$

C. Analysis: Upper Bound

Similar to our analysis in Sections IV-D and V-E, we first upper bound the query complexity in terms of instantenous lower bound on EJS:

Theorem 10. *Consider Problem (P) under Assumptions 3 and 5. Define $\alpha := \min_{\boldsymbol{\rho} \in \mathbb{P}(\Omega)} \max_{x \in \mathcal{X}} EJS(\boldsymbol{\rho}, x)$ which implies that at any given belief vector $\boldsymbol{\rho} \in \mathbb{P}(\Omega)$, it is possible to find a sample $x \in \mathcal{X}$ satisfying $EJS(\boldsymbol{\rho}, x) \geq \alpha$. If $\alpha > 0$, then*

$$\mathbb{E}[\tau_\epsilon^*] \leq \frac{\log M + \max\{\log\log M, \log \frac{1}{\epsilon}\} + 4C_2}{\alpha}. \tag{86}$$

Also define $\beta := \min_{\boldsymbol{\rho} \in \mathbb{P}_\epsilon^M(\Omega)} \max_{x \in \mathcal{X}} EJS(\boldsymbol{\rho}, x)$. If $\beta > \alpha$, then the following bound is obtained

$$\mathbb{E}[\tau_\epsilon^*] \leq \frac{\log M + \max\{\log\log M, \log\log \frac{1}{\epsilon}\}}{\alpha} + \frac{\log \frac{1}{\epsilon}}{\beta} + \frac{3(4C_2)^2}{\alpha\beta}. \tag{87}$$

Next we look at a specific and important function class, termed as locally identifiable and provide nontrivial characterization of α and β, hence, demonstrating the relative looseness/tightness of the upper bounds.

Before we proceed, we provide the following definitions which will allow us to generalize the notion of 1-neighborly, first suggested by [52]; then for this general class, we will obtain non-trivial α and β as defined in Theorem 10. Consider the representation of a pair of samples x and x' in terms of their partitioning of the functions:

Definition 4. A pair of samples $x, x' \in \mathcal{X}$ partition the function class $\mathcal{H}$ in an agreement set $A_{x,x'} := \{h \in \mathcal{H} : h(x) = h(x')\}$ and a disagreement set $\Delta_{x,x'} := \{h \in \mathcal{H} : h(x) \neq h(x')\}$.

Definition 5. A class of functions $\mathcal{H}$ is referred to as locally identifiable if for any $h_i \in \mathcal{H}$, there exist samples $x, x' \in \mathcal{X}$ and labels $l, l' \in \mathcal{L}$ such that either of the following be true

(i) $h_i \in \Delta_{x,x'} \cap H_l^x \cap H_{l'}^{x'}$ and $\mathcal{H} - \{h_i\} = A_{x,x'} \cup (H_{l'}^x \cap H_l^{x'})$, or

(ii) $\{h_i\} = A_{x,x'} \cap H_l^x$ and for all $k \neq l, l'$, $H_k^x \cup H_k^{x'} = \emptyset$.

In essence, the locally identifiable condition implies that for any function $h_i \in \mathcal{H}$, there are (at least) two samples x and x' in $\mathcal{X}$ and two labels l and l' using which h_i can be distinguished from all other functions. In the binary case, this condition is equivalent to $\mathcal{H}$ having teaching dimension of at most 2 [58]. However, for general $\mathcal{L}$, this condition allows for the teaching sets to be tested against two and only two labels l and l'. As we will see in Section VI-C1, local identifiability is a condition that is satisfied by a number of natural function classes.

The performance of $\mathfrak{c}_{EJS}$ when the labeling function class is locally identifiable is characterized by the capacity of the (sub)channel with two inputs $l, l' \in \mathcal{L}$ denoted by $C_{ll'}$, i.e.,

$$C_{ll'} := \min_{g \in \mathbb{P}(\mathcal{Y})} \max\{D(f_l \| g), D(f_{l'} \| g)\}. \tag{88}$$

Also define

$$C_{\min} := \min_{l,l' \in \mathcal{L}, l \neq l'} \min \left\{ C_{ll'}, D\left(f_{l'} \middle\| \frac{1}{2} f_l + \frac{1}{2} f_{l'} \right) \right\}. \tag{89}$$

Proposition 4. *When function class $\mathcal{H}$ is locally identifiable, $\alpha \geq \frac{1}{M} C_{\min}$ and $\beta \geq \tilde{\rho} C_{\min}$. More precisely, for every belief vector $\boldsymbol{\rho}$, there exists an $x \in \mathcal{X}$ such that*

$$EJS(\boldsymbol{\rho}, x) \geq \begin{cases} \frac{1}{M} C_{\min} & \textit{if } \boldsymbol{\rho} \notin \mathbb{P}_\epsilon^M(\Omega) \\ \tilde{\rho} C_{\min} & \textit{otherwise} \end{cases}. \tag{90}$$

The following corollary provides an upper bound on the expected number of sample queries.

Corollary 3. *Consider Problem (P) under Assumptions 3 and 5. If the function class $\mathcal{H}$ is locally identifiable, then*

$$\mathbb{E}[\tau_\epsilon^*] \leq \frac{M(\log M + \max\{\log\log M, \log\log \frac{1}{\epsilon}\})}{C_{\min}} + \frac{\log \frac{1}{\epsilon}}{\tilde{\rho} C_{\min}} + \frac{3M(4C_2)^2}{\tilde{\rho} C_{\min}^2}. \tag{91}$$

Next, we define a subclass of the locally identifiable function class, and show that for this function class, α and β can be selected to match the denominators in the lower bound in (85).

Definition 6. We refer to the function class $\mathcal{H}$ as $\mathcal{R}(\mathcal{H})$-sample-rich for $\mathcal{R}(\mathcal{H}) = \cup_{x \in \mathcal{X}} \Xi^x$. In the special case where $\mathcal{R}(\mathcal{H})$ includes all $(|\mathcal{L}|^M - |\mathcal{L}|)$ non-trivial $|\mathcal{L}|$-partitions of $\mathcal{H}$, we simply refer to $\mathcal{H}$ as sample-rich.

Proposition 5. *When function class $\mathcal{H}$ is sample-rich, $\alpha \geq C$ and $\beta \geq \tilde{\rho} C_1$.*

Proof: To prove Proposition 5, we will show that for all belief vectors $\boldsymbol{\rho}$,

$$\max_{x \in \mathcal{X}} EJS(\boldsymbol{\rho}, x) \geq C,$$

and furthermore,

$$\max_{x \in \mathcal{X}} EJS(\boldsymbol{\rho}, x) \geq \max_{i \in \Omega} \rho_i C_1.$$

Recall from Subsection VI-B that

$$C = \min_{g\in\mathbb{P}(\mathcal{Y})} \max_{l\in\mathcal{L}} D(f_l\|g), \tag{92}$$

and the minimum is achieved by $g^* = \sum_{l\in\mathcal{L}} \pi_l^\star f_l$ where $\pi^\star$ is the capacity achieving input distribution, i.e.,

$$D\bigg(f_k\bigg\|\sum_{l\in\mathcal{L}}\pi_l^\star f_l\bigg) = C \quad \text{for any } k\in\mathcal{L} \text{ such that } \pi_k^\star > 0. \tag{93}$$

By definition of the sample-rich function class, for each $\boldsymbol{v} := [v_1,\ldots,v_M] \in \mathcal{L}^M$, there exists a sample in $\mathcal{X}$, say $x_{\boldsymbol{v}}$, that satisfies $\boldsymbol{h}(x_{\boldsymbol{v}}) = \boldsymbol{v}$, where $\boldsymbol{h}(x) := [h_1(x), h_2(x),\ldots,h_M(x)]$. Let

$$\lambda_{\boldsymbol{v}}^\star = \prod_{i=1}^M \pi_{v_i}^\star.$$

Note that $\sum_{\boldsymbol{v}\in\mathcal{L}^M}\lambda_{\boldsymbol{v}}^\star = 1$. Moreover, for any $i,j\in\Omega$, $i\neq j$,

$$\sum_{\boldsymbol{v}\in\mathcal{L}^M:\, v_i=k}\lambda_{\boldsymbol{v}}^\star = \pi_k^\star, \qquad \sum_{\boldsymbol{v}\in\mathcal{L}^M:\, v_i=k, v_j=l}\lambda_{\boldsymbol{v}}^\star = \pi_k^\star\pi_l^\star.$$

Using weights $\{\lambda_{\boldsymbol{v}}^\star\}_{\boldsymbol{v}\in\mathcal{L}^M}$ and taking average over all $\boldsymbol{v}\in\mathcal{L}^M$, we obtain

$$\begin{aligned}
&\max_{x\in\mathcal{X}} EJS(\boldsymbol{\rho},x)\\
&\geq \sum_{\boldsymbol{v}}\lambda_{\boldsymbol{v}}^\star EJS(\boldsymbol{\rho},x_{\boldsymbol{v}})\\
&= \sum_{\boldsymbol{v}}\lambda_{\boldsymbol{v}}^\star\sum_{i=1}^M\rho_i D\bigg(f_{h_i(x_{\boldsymbol{v}})}\bigg\|\sum_{j\neq i}\frac{\rho_j}{1-\rho_i}f_{h_j(x_{\boldsymbol{v}})}\bigg)\\
&= \sum_{i=1}^M\rho_i\sum_{k\in\mathcal{L}}\pi_k^\star\sum_{\boldsymbol{v}:\, v_i=k}\frac{\lambda_{\boldsymbol{v}}^\star}{\pi_k^\star}D\bigg(f_k\bigg\|\sum_{j\neq i}\frac{\rho_j}{1-\rho_i}f_{v_j}\bigg)\\
&\overset{(a)}{\geq} \sum_{i=1}^M\rho_i\sum_{k\in\mathcal{L}}\pi_k^\star D\bigg(f_k\bigg\|\sum_{j\neq i}\frac{\rho_j}{1-\rho_i}\sum_{\boldsymbol{v}:\, v_i=k}\frac{\lambda_{\boldsymbol{v}}^\star}{\pi_k^\star}f_{v_j}\bigg)\\
&= \sum_{i=1}^M\rho_i\sum_{k\in\mathcal{L}}\pi_k^\star D\bigg(f_k\bigg\|\sum_{j\neq i}\frac{\rho_j}{1-\rho_i}\sum_{l\in\mathcal{L}}\sum_{\boldsymbol{v}:\, v_i=k,v_j=l}\frac{\lambda_{\boldsymbol{v}}^\star}{\pi_k^\star}f_l\bigg)\\
&= \sum_{i=1}^M\rho_i\sum_{k\in\mathcal{L}}\pi_k^\star D\bigg(f_k\bigg\|\sum_{l\in\mathcal{L}}\pi_l^\star f_l\bigg)\\
&\overset{(b)}{=} \sum_{i=1}^M\rho_i C\\
&= C,
\end{aligned}$$

where (a) follows from Jensen's inequality and (b) follows from (93).

Let $\hat{i} = \arg\max_{i\in\Omega}\rho_i$. Let $k,l\in\mathcal{L}$ be the labels satisfying $D(f_k\|f_l) = C_1$. By definition of the sample-rich function class, there exists a sample $x_{\hat{i}}\in\mathcal{X}$ that satisfies $h_{\hat{i}}(x_{\hat{i}}) = k$ and $h_j(x_{\hat{i}}) = l$ for all $j\neq\hat{i}$. We have

$$\begin{aligned}
\max_{x\in\mathcal{X}} EJS(\boldsymbol{\rho},x) &\geq EJS(\boldsymbol{\rho},x_{\hat{i}})\\
&\geq \rho_{\hat{i}} D\bigg(f_{h_{\hat{i}}(x_{\hat{i}})}\bigg\|\sum_{j\neq\hat{i}}\frac{\rho_j}{1-\rho_{\hat{i}}}f_{h_j(x_{\hat{i}})}\bigg)\\
&= \max_{i\in\Omega}\rho_i C_1.
\end{aligned}$$

■

As a simple corollary,

Corollary 4. *Consider Problem (P) under Assumptions 3 and 5. If the function class $\mathcal{H}$ is sample-rich,*

$$\mathbb{E}[\tau_\epsilon^*] \leq \frac{\log M + \max\{\log\log M, \log\log\frac{1}{\epsilon}\}}{C} + \frac{\log\frac{1}{\epsilon}}{\tilde{\rho}C_1} + \frac{48C_2^2}{\tilde{\rho}CC_1}. \tag{94}$$

The above results show that for sample-rich function classes, the lower and upper bounds are tight (their dominant terms in both ϵ and M match).

The above results generalize the finding of [52] to a multi-label Bayesian learning with non-binary and asymmetric noise case. However, to make this comparison precise, we will dedicate Section VI-C1 to specialize our general results above to the noisy generalized binary search of [52].

1) Special Case: Noisy Generalized Binary Search: We next compare our work with existing results. Since the only study of similar nature is that of noisy generalized binary search [52], we consider an application of our main results to noisy generalized binary search among 1-neighborly functions, first introduced in [52]. This is a special case of our problem where functions are binary-valued, i.e., $\mathcal{L} = \{-1,+1\}$, the observation space $\mathcal{Y} = \{-1,+1\}$, and observation distributions are of the following form:

$$f_l(y) = \begin{cases} 1-p & \text{if } y = l\\ p & \text{if } y = -l\end{cases},$$

for some $p\in(0,1/2)$. In other words, for any sample x, if h_i is the true function, then the label $h_i(x)$ is observed through a binary symmetric channel with crossover probability p.

For the case of noisy generalized binary search, C, C_1, and C_2 defined in Subsection VI-B can be further simplified to

$$\begin{aligned}
C &:= 1 + p\log p + (1-p)\log(1-p),\\
C_1 &:= p\log\frac{p}{1-p} + (1-p)\log\frac{1-p}{p},\\
C_2 &:= \frac{1-p}{p}.
\end{aligned}$$

In order to emphasize the dependence of C, C_1, and C_2 on the Bernoulli parameter p (corresponding to the observation noise), we denote them by $C(p)$, $C_1(p)$, and $C_2(p)$ respectively. Note that from Jensen's inequality, $C_1(p)\geq 2C(p)$.

Next we define a class of 1-neighborly functions first defined in [52, Definition 2].

Definition 7. Two samples x, x' are said to be 1-neighbor if only a single function (and its complement, if it also belongs to $\mathcal{H}$) outputs a different value on x and x'. A class of binary-valued functions $\mathcal{H}$ is referred to as 1-neighborly if the 1-neighborhood graph of $\mathcal{X}$ is connected, i.e., for every pair of samples in $\mathcal{X}$ there exists a sequence of 1-neighbor samples that begins at one of the pair and ends with the other.

It is simple to see that the class of 1-neighborly functions is a subset of binary-valued locally identifiable function class. (Remember that in the binary case, the locally identifiable condition is equivalent to $\mathcal{H}$ having teaching dimension of at most 2, i.e., any function $h_i \in \mathcal{H}$ can be uniquely identified using one or at most two samples from $\mathcal{X}$).

This implies the following baseline bound:

Corollary 5. *When function class $\mathcal{H}$ is 1-neighborly, we have $\alpha \geq \frac{1}{M}C(p)$ and $\beta \geq \tilde{\rho}C(p)$.*

Remark 7. The statement of Corollary 5 continues to hold for any function class $\mathcal{H}$ which has teaching dimension of at most 2. Although this class of functions slightly generalizes the 1-neighborly class, in this section we focus on the class of 1-neighborly functions and its specific subclasses in order to allow for comparison with previous studies [52].

In comparison, [52] provides two sample query strategies, NGBS and MSGBS, whose performance (upper bound) depends strongly on the properties of the function class at hand.

Let n_0 denote the number of queries made by GBS to determine h_θ in the noiseless setting. The number of queries required by NGBS to attain $\mathrm{Pe} \leq \epsilon$ is upper bounded by

$$\frac{n_0(\log n_0 + \log \frac{1}{\epsilon})}{(\frac{1}{2} - p)^2}. \tag{95}$$

Let $\mathcal{A}$ denote the smallest partition of sample space $\mathcal{X}$, i.e., $\mathcal{X} = \cup_{A \in \mathcal{A}} A$, such that for every $A \in \mathcal{A}$ and $h \in \mathcal{H}$, the value of $h(x)$ is constant for all $x \in A$; and denote this value by $h(A)$. Furthermore, let

$$c^* := \min_{P \in \mathbb{P}(\mathcal{A})} \max_{h \in \mathcal{H}} \left| \sum_{A \in \mathcal{A}} h(A)P(A) \right|. \tag{96}$$

For a 1-neighborly function class of size M, the following upper bound for n_0 is obtained[9]

$$n_0 \leq \left\lceil \frac{\log M}{-\log\left(\max\left\{\frac{1+c^*}{2}, \frac{2}{3}\right\}\right)} \right\rceil. \tag{97}$$

On the other hand, under MSGBS, the number of queries required to ensure that $\mathrm{Pe} \leq \epsilon$ for a 1-neighborly function class of size M is upper bounded by

$$\frac{\log M + \log \frac{1}{\epsilon}}{\min\{2(1-c^*), 1\}\lambda(p)}, \tag{98}$$

where

$$\lambda(p) := \max_{p' \in (p, 1/2)} \frac{\log e}{4} \left(1 - \frac{p'(1-p)}{1-p'} - \frac{(1-p')p}{p'}\right). \tag{99}$$

Note that c^* (as well as n_0) in general depends on the function class $\mathcal{H}$. Since this dependence is implicit and hard to characterize in closed form for general function class $\mathcal{H}$, a direct comparison between (98) (or (95)) and (91) is not possible. As a result, next we focus on special cases of function classes studied in [52] for which a precise characterization of the achievable upper bound is available. Consequently, we next define two important subclasses of 1-neighborly binary-valued functions: 1) Disjoint class $\mathcal{H}_D$; 2) Threshold class $\mathcal{H}_T$. We further specialize α and β for these classes.

Definition 8. Let $\boldsymbol{e}_i$, $i \in \Omega$, represent a vector of size M whose i^{th} element is $+1$ and all other elements are -1. A collection of functions $\mathcal{H}$ is referred to as *disjoint interval class* if $\cup_{x \in \mathcal{X}}\{\boldsymbol{h}(x)\} = \cup_{i \in \Omega}\{\boldsymbol{e}_i\} \subset \{-1,+1\}^M$, where $\boldsymbol{h}(x) = [h_1(x), h_2(x), \ldots, h_M(x)]$. In other words, for any sample $x \in \mathcal{X}$, only one function in $\mathcal{H}$ takes value $+1$ and all other functions take value -1.

Definition 9. Let $\boldsymbol{u}_i$, $i \in \Omega$, represent a vector of size M whose first i elements are -1 and all other elements are $+1$. A collection of functions $\mathcal{H}$ is referred to as *threshold class* if $\cup_{x \in \mathcal{X}}\{\boldsymbol{h}(x)\} = \cup_{i \in \Omega}\{\boldsymbol{u}_i\} \subset \{-1,+1\}^M$.

Fact 6 (see [20]). *For the disjoint interval class $\mathcal{H}_D$, $c^* = 1 - \frac{2}{M}$ and $n_0 \leq M$. For the threshold function class $\mathcal{H}_T$, $c^* = 0$ and $n_0 \leq \left\lceil \frac{\log M}{\log \frac{3}{2}} \right\rceil \leq 2\log M$. For the sample-rich function class $\mathcal{H}_R$, $c^* = 0$ and $n_0 \leq \left\lceil \frac{\log M}{\log \frac{3}{2}} \right\rceil \leq 2\log M$.*

We are now ready to contrast these results with our findings. In particular, we have

Proposition 6. *For the disjoint interval class $\mathcal{H}_D$, $\alpha \geq \frac{1}{M}C_1(p)$ and $\beta \geq \tilde{\rho}C_1(p)$. For the threshold function class $\mathcal{H}_T$, $\alpha \geq C(p)$ and $\beta \geq C(p)$. For the sample-rich function class $\mathcal{H}_R$, $\alpha \geq C(p)$ and $\beta \geq \tilde{\rho}C_1(p)$.*

D. Comparison to Known Results

In this section, we compare the performance of $\mathfrak{c}_{EJS}$ to that of NGBS and MSGBS policies proposed in [20].

Table I summarizes our results and specializes the upper bounds in [20] and lists the number of samples required by the policies NGBS, MSGBS, and $\mathfrak{c}_{EJS}$ to attain $\mathrm{Pe} \leq \epsilon$. Furthermore, these bounds together with the lower bound provide a sense as the relative tightness of the bounds for small ϵ and/or large M.[10]

Recall that policies NGBS and MSGBS are non-sequential in the sense that they stop after a fixed number of samples, regardless of the probability of error. The numbers shown in Table I are the number of samples that these policies

[9] In general, the GBS algorithm is able to determine the true function with label complexity at most $\lceil \log M \rceil$ times the extended teaching dimension [59]. Moreover, in the case of 1-neighborly classes, [60, Lemma 8.6] has shown that the extended teaching dimension is dominated by the two constant classifiers.

[10] The term $o(1)$ in the bounds goes to zero as $\epsilon \to 0$ or $M \to \infty$. See [2] for more details.

TABLE I

PERFORMANCE COMPARISON OF NGBS, MSGBS, AND $\mathfrak{c}_{EJS}$ ON DIFFERENT FUNCTION CLASSES.

Function class	NGBS	MSGBS	$\mathfrak{c}_{EJS}$
Disjoint $\mathcal{H}_D$	$\frac{M(\log M+\log\frac{1}{\epsilon})}{(\frac{1}{2}-p)^2}$	$\frac{M(\log M+\log\frac{1}{\epsilon})}{4\lambda(p)}$	$\left(\frac{M\log M}{C_1(p)}+\frac{\log\frac{1}{\epsilon}}{C_1(p)}\right)(1+o(1))$
Threshold $\mathcal{H}_T$	$\frac{2\log M(\log\log M+\log\frac{1}{\epsilon}+1)}{(\frac{1}{2}-p)^2}$	$\frac{\log M+\log\frac{1}{\epsilon}}{\lambda(p)}$	$\left(\frac{\log M}{C(p)}+\frac{\log\frac{1}{\epsilon}}{C(p)}\right)(1+o(1))$
Sample-rich $\mathcal{H}_R$	$\frac{2\log M(\log\log M+\log\frac{1}{\epsilon}+1)}{(\frac{1}{2}-p)^2}$	$\frac{\log M+\log\frac{1}{\epsilon}}{\lambda(p)}$	$\left(\frac{\log M}{C(p)}+\frac{\log\frac{1}{\epsilon}}{C_1(p)}\right)(1+o(1))$
Lower Bound on the number of samples needed		$\geq\left(\frac{\log M}{C(p)}+\frac{\log\frac{1}{\epsilon}}{C_1(p)}\right)(1-o(1))$	

require to achieve $\mathrm{Pe} \leq \epsilon$. Policy $\mathfrak{c}_{EJS}$ is sequential and Table I shows the expected number of samples required by this policy to achieve $\mathrm{Pe} \leq \epsilon$.

Finally, to provide a comparison between the obtained bounds, in asymptotic regime, Fig. 4 compares the denominators of the upper bounds given in Table I. Note that our upper bound provides improvement over those corresponding to NGBS and MSGBS. Particularly, the gap between the bounds is very significant for small values of the Bernoulli parameter p and for large values of $\frac{1}{\epsilon}$ and M.

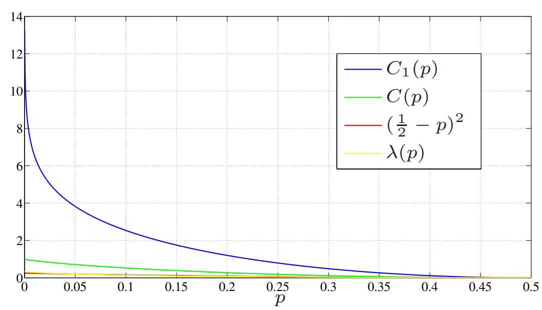

Fig. 4. Comparison of $C(p)$, $C_1(p)$, $(\frac{1}{2}-p)^2$, and $\lambda(p)$, for $p \in (0, 1/2)$.

VII. DISCUSSION AND FUTURE WORK

This paper focuses on the problem of information acquisition and utilization where a decision maker dynamically refines his/her belief about stochastically time-varying parameters in order to utilize a system of interest as efficiently as possible. A new theoretical framework for stochastic learning and decision-making in such a setting termed *Information Acquisition and Utilization Problems* was proposed. IAUP is a special case of partially observable Markov decision problems (POMDP) with several unique properties, with the most significant one is the independence of the underlying stochastic process describing the time-varying parameters of interest and the decision maker's selection of of the acquisition and utilization actions.

Motivated by a synthesis of the prior works on active hypothesis testing, noisy dynamic search, and noisy Bayesian active learning, the IAUP framework borrows from diverse areas of research from statistics, information theory, and stochastic control. This framework, naturally, motivates a set ongoing and future research questions both in the domain of algorithm design as well as analytic/tight performance lower bounds on the minimum feasible cost.

REFERENCES

[1] D.P. Bertsekas, *Dynamic Programming and Optimal Control, Volume II*, Athena Scientific, 1995.

[2] M. Naghshvar, "Active learning and hypothesis testing," 2013, PhD Thesis, University of California San Diego.

[3] Yonatan Kaspi, Ofer Shayevitz, and Tara Javidi, "Searching with measurement dependent noise," *2014 IEEE Information Theory Workshop, ITW 2014*, pp. 267–271, 2014.

[4] M. Naghshvar, T. Javidi, and K. Chaudhuri, "Bayesian active learning with non-persistent noise," *IEEE Transactions on Information Theory*, vol. 61, no. 7, pp. 4080–4098, July 2015.

[5] T. Javidi and A. Goldsmith, "Dynamic joint source–channel coding with feedback," in *IEEE International Symposium on Information Theory (ISIT)*, 2013.

[6] D. P. Bertsekas and S. E. Shreve, *Stochastic Optimal Control*, Athena Scientific, Belmont, Massachusetts, 1996.

[7] L. K. Platzman, "Optimal infinite-horizon undiscounted control of finite probabilistic systems," *SIAM Journal of Control And Optimization*, vol. 18, July 1980.

[8] D. P. Bertsekas and S. E. Shreve, *Stochastic optimal control: The discrete-time case*, Athena Scientific, Belmont, MA, 2007.

[9] D. Blackwell, "Equivalent Comparisons of Experiments," *The Annals of Mathematical Statistics*, vol. 24, pp. 265–272, 1953.

[10] M. H. DeGroot, *Optimal Statistical Decisions*, McGraw-Hill, Inc., 1970.

[11] H. Jeffreys, "An invariant form for the prior probability in estimation problems," *Proceedings of the Royal Society. London. Series A.*, vol. 186, pp. 453–461, 1946.

[12] J. Lin, "Divergence Measures Based on the Shannon Entropy," *IEEE Transactions on Information Theory*, vol. 37, no. 1, pp. 145 –151, Jan 1991.

[13] J. Burbea and C. R. Rao, "On the convexity of some divergence measures based on entropy functions," *IEEE Transactions on Information Theory*, vol. 28, no. 3, pp. 489–495, May 1982.

[14] G. T. Toussaint, "Some functional lower bounds on the expected divergence for multihypothesis pattern recognition, communication, and radar systems," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-1, pp. 384–385, 1971.

[15] M. Naghshvar and T. Javidi, "Active Sequential Hypothesis Testing," *Annals of Statistics*, 2013, available on arXiv:1203.4626.

[16] S. M. Berry, B. P. Carlin, J. J. Lee , and P. Müller, *Bayesian Adaptive Methods for Clinical Trials*, CRC Press, 2010.

[17] P. Shenoy and A. J. Yu, "Rational Decision-making in Inhibitory Control," *Frontiers in Human Neuroscience*, vol. 5, no. 48, 2011.

[18] A. O. Hero and D. Cochran, "Sensor Management: Past, Present, and Future," *IEEE Sensors Journal*, vol. 11, no. 12, pp. 3064 –3075, December 2011.

[19] G. A. Hollinger, U. Mitra, and G. S. Sukhatme, "Active classification: Theory and application to underwater inspection," in *International Symposium on Robotics Research*, 2011.

[20] R. D. Nowak, "The geometry of generalized binary search," *IEEE Transactions on Information Theory*, vol. 57, no. 12, pp. 7893–7906, December 2011.

[21] C. L. Chan, P. H. Che, S. Jaggi, and V. Saligrama, "Non-adaptive probabilistic group testing with noisy measurements: near-optimal bounds with efficient algorithms," in *49th Annual Allerton Conference on Communication, Control, and Computing*, 2011, pp. 1832–1839.

[22] M. V. Burnashev, "Data Transmission Over a Discrete Channel with Feedback Random Transmission Time," *Problemy Peredachi Informatsii*, vol. 12, no. 4, pp. 10–30, 1975.

[23] A. Wald and J. Wolfowitz, "Optimal Character of the Sequential Probability Ratio Tests," *The Annals of Mathematical Statistics*, vol. 19, no. 3, pp. 326–339, 1948.

[24] P. Armitage, "Sequential Analysis with more than Two Alternative Hypotheses, and its Relation to Discriminant Function Analysis," *Journal of the Royal Statistical Society, Series B*, vol. 12, no. 1, pp. 137–144, 1950.

[25] G. Lorden, "Nearly-optimal sequential tests for finitely many parameter values," *The Annals of Statistics*, vol. 5, no. 1, pp. 1–21, 1977.

[26] V. P. Dragalin, A. G. Tartakovsky, and V. V. Veeravalli, "Multihypothesis Sequential Probability Ratio Tests. I: Asymptotic Optimality," *IEEE Transactions on Information Theory*, vol. 45, no. 7, pp. 2448–2461, November 1999.

[27] D. V. Lindley, "On a Measure of the Information Provided by an Experiment," *The Annals of Mathematical Statistics*, vol. 27, no. 4, pp. 986–1005, 1956.

[28] L. Le. Cam, "Sufficiency and Approximate Sufficiency," *The Annals of Mathematical Statistics*, vol. 35, no. 4, pp. 1419–1455, 1964.

[29] K. Goel and M. H. DeGroot, "Comparison of Experiments and Information Measures," *The Annals of Statistics*, vol. 7, no. 5, pp. 1066–1077, 1979.

[30] E. L. Lehmann, "Comparing Location Experiments," *The Annals of Statistics*, vol. 16, no. 2, pp. 521–533, 1988.

[31] E. Torgersen, *Stochastic Orders and Comparison of Experiments*, vol. 19, pp. 334–371, Hayward, CA: Institute of Mathematical Statistics, 1991.

[32] M. Naghshvar and T. Javidi, "Active M-ary Sequential Hypothesis Testing," in *IEEE International Symposium on Information Theory (ISIT)*, 2010, pp. 1623–1627.

[33] M. H. DeGroot, "Uncertainty, information, and sequential experiments," *The Annals of Mathematical Statistics*, vol. 33, no. 2, pp. 404–419, 1962.

[34] H. Chernoff, "Sequential Design of Experiments," *The Annals of Mathematical Statistics*, vol. 30, pp. 755–770, 1959.

[35] S. A. Bessler, "Theory and applications of the sequential design of experiments, K-Actions and infinitely many experiments: Part I — Theory," *Technical Report no. 55, Department of Statistics, Stanford University*, 1960.

[36] A. E. Albert, "The sequential design of experiments for infinitely many states of nature," *The Annals of Mathematical Statistics*, vol. 32, pp. 774–799, 1961.

[37] J. Kiefer and J. Sacks, "Asymptotically optimum sequential inference and design," *The Annals of Mathematical Statistics*, vol. 34, no. 3, pp. 705–750, 1963.

[38] W. J. Blot and D. A. Meeter, "Sequential experimental design procedures," *Journal of the American Statistical Association*, vol. 68, no. 343, pp. 586–593, 1973.

[39] R. Keener, "Second order efficiency in the sequential design of experiments," *The Annals of Statistics*, vol. 12, no. 2, pp. 510–532, 1984.

[40] S. P. Lalley and G. Lorden, "A control problem arising in the sequential design of experiments," *The Annals of Probability*, vol. 14, no. 1, pp. 136–172, 1986.

[41] S. Nitinawarat, G. Atia, and V. V. Veeravalli, "Controlled sensing for multihypothesis testing," 2012, available on arXiv:1205.0858v4.

[42] C. E. Shannon, "The Zero Error Capacity of a Noisy Channel," *IRE Transactions on Information Theory*, vol. 2, pp. 8–19, 1956.

[43] M. V. Burnashev, "Sequential Discrimination of Hypotheses with Control of Observations," *Math. USSR Izvestija*, vol. 15, no. 3, pp. 419–440, 1980.

[44] Marat Valievich Burnashev and Kamil'Shamil'evich Zigangirov, "An interval estimation problem for controlled observations," *Problemy Peredachi Informatsii*, vol. 10, no. 3, pp. 51–61, 1974.

[45] C. E. Shannon, "The zero-error capacity of a noisy channel," *IRE. Trans. Info. Theory*, vol. IT-2, pp. 8–19, Sept 1956.

[46] M. Horstein, "Sequential Transmission Using Noiseless Feedback," *IEEE Transactions on Information Theory*, vol. 9, no. 3, pp. 136–143, July 1963.

[47] H. Yamamoto, "Rate-distortion theory for the shannon cipher system," *IEEE Transactions on Information Theory*, vol. 43, no. 3, pp. 827–835, May 1997.

[48] Mohammad Naghshvar and Tara Javidi, "Two-Dimensional Visual Search," *IEEE International Symposium on Information Theory*, pp. 1262–1266, 2013.

[49] Mohammad Naghshvar, Tara Javidi, and Michèle Wigger, "Extrinsic Jensen-Shannon divergence: Applications to variable-length coding," *IEEE Transactions on Information Theory*, vol. 61, no. 4, pp. 2148–2164, 2015.

[50] H. Yamamoto and K. Itoh, "Asymptotic performance of a modified schalkwijk-barron scheme for channels with noiseless feedback (corresp.)," *Information Theory, IEEE Transactions on*, vol. 25, no. 6, pp. 729–733, Nov 1979.

[51] O. Shayevitz and M. Feder, "Optimal feedback communication via posterior matching," *IEEE Transactions on Information Theory*, vol. 57, no. 3, pp. 1186–1222, March 2011.

[52] R. D. Nowak, "Noisy Generalized Binary Search," in *NIPS*, 2009.

[53] M. Naghshvar and T. Javidi, "Extrinsic Jensen–Shnnon Divergence with Application in Active Hypothesis Testing," in *IEEE International Symposium on Information Theory (ISIT)*, 2012.

[54] T. M. Cover and J. A. Thomas, *Elements of information theory*, John Wiley & Sons, Inc., Hoboken, NJ, second edition, 2006.

[55] R. G. Gallager, *Information Theory and Reliable Communication*, John Wiley & Sons, Inc., New York, 1968.

[56] P. Berlin, B. Nakiboglu, B. Rimoldi, and E. Telatar, "A Simple Converse of Burnashev's Reliability Function," *IEEE Transactions on Information Theory*, vol. 55, pp. 3074–3080, 2009.

[57] Y. Polyanskiy, H. V. Poor, and S. Verdu, "Feedback in the Non-Asymptotic Regime," *IEEE Transactions on Information Theory*, vol. 57, no. 8, pp. 4903–4925, Aug 2011.

[58] Sally A. Goldman and Michael J. Kearns, "On the complexity of teaching," *J. Comput. Syst. Sci.*, vol. 50, no. 1, pp. 20–31, 1995.

[59] José L. Balcázar, Jorge Castro, and David Guijarro, "A new abstract combinatorial dimension for exact learning via queries," *J. Comput. Syst. Sci.*, vol. 64, no. 1, pp. 2–21, 2002.

[60] Steve Hanneke, *Theory of Active Learning (Version 1.1)*, 2014.

APPENDIX

The main building block of our analysis is the following concentration inequality established in [2].

Fact 7 (Lemma 4.7.4 in [2])**.** *Assume that the sequence* $\{\xi(t)\}$, $t = 0, 1, 2, \ldots$ *forms a submartingale with respect to a filtration* $\{\mathcal{F}(t)\}$ *and positive constants* K_1, K_2, *and* K_3*:*

$$\begin{aligned}
&\mathbb{E}[\xi(t+1)|\mathcal{F}(t)] \geq \xi(t) + K_1 \quad \textit{if} \quad \xi(t) < 0, \\
&\mathbb{E}[\xi(t+1)|\mathcal{F}(t)] \geq \xi(t) + K_2 \quad \textit{if} \quad \xi(t) \geq 0, \\
&|\xi(t+1) - \xi(t)| \leq K_3 \quad \textit{if} \quad \max\{\xi(t+1), \xi(t)\} \geq 0.
\end{aligned}$$

Consider the stopping time $\upsilon = \min\{t : \xi(t) \geq B\}$*,* $B > 0$*. Then we have the inequality*

$$\mathbb{E}[\upsilon] \leq \frac{B - \xi(0)}{K_2} + \xi(0)\mathbf{1}_{\{\xi(0)<0\}}\left(\frac{1}{K_2} - \frac{1}{K_1}\right) + \frac{3K_3^2}{K_1K_2}.$$