Discrete State Estimation with Persistent Sensor Faults and Non-Persistent Noise via Noisy Bayesian Active Diagnosis

Sze Zheng Yong Necmiye Ozay

***Abstract*— In this paper, we consider adaptive decision-making problems for stochastic discrete state estimation with a given budget of sensing actions/measurements that yield noisy and/or faulty partial observations. This problem is an extension of Bayesian active diagnosis, which is known to be NP-hard, to the setting when the sensor measurements are vector-valued and may be affected by persistent sensor faults and/or non-persistent noise. In particular, we identify meaningful reward functions for this problem that are *adaptive monotone* and *weakly adaptive submodular*; thus an adaptive greedy algorithm (with no need for proxy reward functions nor new algorithms) has guaranteed near-optimal performance. Finally, we apply our approach to discrete state estimation via active sensing of an electrical power system with sensor faults (*persistent* noise) and sensor noise (stochastic/*non-persistent* noise).**

I. Introduction

The operation of complex cyber-physical systems often involves a sequence of control and sensing decisions under partial observability. This problem of sequential decision-making appears in various applications, both in the context of stochastic control (e.g., in robot navigation [1], reinforcement learning [2]) and stochastic state estimation (e.g., in information gathering in robotics [3], [4], fault diagnosis in nuclear plants [5], sensor placement and scheduling [6]–[9]). Specifically, the research area of active diagnosis and active learning, i.e., the problem of discrete state (or diagnosis or hypothesis) identification via a minimal number of sequential information-bearing sensing actions, has many real-world applications in medical diagnosis and emergency response [10], [11], electrical systems diagnosis [12], [13], etc., and thus can have significant and broad impacts on the fields of cyber-physical systems, robotics and artificial intelligence.

Related Work. Finding optimal active sensing policies for general partially observable stochastic estimation problems is often an intractable combinatorial optimization problem. Hence, researchers, e.g., [8], [14], [15] in the context of actuator and sensor placement and scheduling, have appealed to a diminishing returns property known as submodularity, which plays a similar role in combinatorial optimization as convexity in continuous optimization. Furthermore, a recent paper [6] introduced the notion of adaptive submodularity for set functions that extends submodularity to the adaptive setting, i.e., when actions are sequentially chosen based on observations/sensor measurements, and provided near-optimal performance guarantees for adaptive greedy policies.

When the sensor measurements or observations are not corrupted by noise or faults, [6] identified a reward function for the active learning and diagnosis problems that they showed to be adaptive monotone submodular and further proved that an adaptive greedy policy yields the best polynomial-time approximation algorithm, based on the results of [16]. The setting when the observations are corrupted by noise or faults has also been studied, but only for special cases, of either only persistent faults or only non-persistent noise, although they can occur simultaneously in many applications with vector-valued observations, as is considered here.

When the noise is stochastic/non-persistent, i.e., the case when repeated measurements may yield different observations, [17], [18] proposed practical algorithms based on rank and entropy approximation, respectively, for active diagnosis without providing performance guarantees. For the complementary problem of (non-persistently) noisy active learning that selects sensing actions to identify the correct state/diagnosis with a given maximum probability of error, noisy generalized binary search (GBS) and Extrinsic Jensen-Shannon (EJS) divergence maximization have been proposed in [19], [20], respectively, each with its own performance guarantees. These approaches do not rely on (weak) adaptive submodularity [6], [13] that can potentially simplify the performance analysis.

On the other hand, the active diagnosis problem with persistent noise or faults has only been recently considered in [13], although the complementary Bayesian active learning with persistent noise has been investigated earlier in [4], [11], [21]. All these approaches consider a more general problem of group-based active diagnosis and learning via defining the possible states/diagnoses of interest as groups of modes/objects of persistent faults. [4], [11], [21] share the indirect approach of defining proxy reward functions that are adaptive submodular, providing new corresponding algorithms and consequently proving performance guarantees for the original problem. In contrast, the direct approach in [13] showed that the reward function is *weakly* adaptive submodular and proved that the adaptive greedy algorithm for the unmodified reward function also has similar performance guarantees as [11], [21].

Contributions. We consider the noisy discrete state estimation problem (more generally, the noisy Bayesian active diagnosis problem) where the vector-valued observations

This work was supported in part by an Early Career Faculty grant from NASA's Space Technology Research Grants Program and DARPA grant N66001-14-1-4045.

S.Z. Yong is with the School for Engineering of Matter, Transport and Energy, Arizona State University, Tempe, AZ, USA (e-mail: szyong@asu.edu).

N. Ozay is with the Department of Electrical and Computer Engineering, University of Michigan, Ann Arbor, MI, USA (e-mail: necmiye@umich.edu).

may be corrupted by persistent sensor faults and/or non-persistent stochastic noise. We propose reward functions that are meaningful for robust decision-making/control problems in safety-critical and health-related applications, which we further show to be *adaptive monotone* and *weakly adaptive submodular*; hence, an adaptive greedy active sensing policy (with no need for sophisticated new algorithms) has guaranteed near-optimal performance. When we only have persistent sensor faults or when there are no faults or noise, the reward functions reduce to the ones in the literature [6], [13] and when the noise is only non-persistent, the proposed greedy policies provide alternatives to the approaches in [19], [20]. Lastly, we empirically evaluate our approach with the proposed reward functions on a state estimation problem of an aircraft electrical system, which we observed to perform just as well as an exhaustive search policy while using significantly less computation.

II. Motivation: Aircraft Electrical System with Persistent Faults and Stochastic Noise

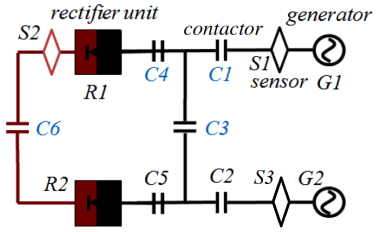

Fig. 1: A single-line diagram of a simple circuit with AC components (in black) and DC components (in red).

We are motivated by the discrete state estimation problem of an aircraft electrical system via active sensing, first studied in [12] (see Figure 1 for an example of a simple circuit and the readers are referred to [12] for a detailed description of the electrical circuits). As more systems become more dependent on electric power systems, this problem is crucial to the operation and safety of these systems. In [12], the sensors are assumed to be *healthy* or faultless and noiseless, while [13] considers the setting with persistent sensor faults that persistently provide incorrect information about the unknown discrete state (i.e., operating condition or status) of the electrical components. To complement [13], we will also consider non-persistent or stochastic sensor noise in this paper. In contrast to persistent faults, a non-persistently noisy sensor outputs a false measurement with a certain probability each time a measurement/sensing action is taken, and thus, the sensor measurements do not stay constant/persistent when the same sensing action is repeated. Thus, in this case, there is a trade-off between exploration (taking a separate sensing action) and exploitation (repeating the same action).

Despite the presence of persistent sensor faults and/or non-persistent sensor noise, our goal is similar to that in [12], [13], i.e., to design a policy that can adaptively/sequentially estimate the discrete state of the circuit by taking active sensing "actions" (i.e., opening or closing controllable contactors) and observing the sensor measurements.

III. Problem Formulation

We consider the discrete state estimation problem (also known as Bayesian active diagnosis or adaptive stochastic maximization) with a finite set of vector-valued states/hypotheses/diagnoses $\mathcal{X} = \mathcal{X}_1 \times \ldots \times \mathcal{X}_n$ (e.g., health of electrical components, locations of target objects) with a (possibly non-uniform) prior probability distribution $\mathbb{P}[x]$, a finite set of sensing actions/queries/tests $\mathcal{V}$ and a finite set of vector-valued sensor measurements/observations $\mathcal{Y} = \mathcal{Y}_1 \times \ldots \times \mathcal{Y}_m$ (e.g., an array of sensors, multi-question questionnaire). In contrast to the traditional noiseless discrete state estimation problem, the sensing actions can yield noisy observations, which can be persistent and/or non-persistent. If the noise affecting observation $y_i \in \mathcal{Y}_i$ is persistent (henceforth referred to as deterministic/persistent fault), then repeating the same action will yield the same observation, whereas if the noise affecting observation $y_i \in \mathcal{Y}_i$ is non-persistent (henceforth referred to as stochastic noise), then repeated actions may yield different observations. Note that the stochastic noise should not be viewed as a special case of the persistent fault with an infinite number of copies of each action (a common misconception) because each repeated action would incur a cost. Moreover, since each action (repeated or not) incurs a cost, there is a net loss in repeating an action if we only have persistent sensor faults but the repeated action may result in a net gain if there is also stochastic or non-persistent sensor noise, leading to a form of trade-off between exploration and exploitation.

Deterministic faults are modeled as *fault modes* that do not change with time, denoted $q \in \mathcal{Q}$, and is a $\mathcal{Q}$-valued random variable with conditional probability mass function $\mathbb{P}[q|x]$. We will also adopt the types of faults in [13], where a Type 1 fault is when a faulty observation persistently outputs another (wrong) outcome and a Type 2 fault is when the faulty observation is always constant, e.g., '0' or '1'. On the other hand, stochastic noise are time-dependent *noise modes*, i.e., $w_t \in \mathcal{W}$ for each time t that an action/test is chosen. The noise mode for each time t is determined by a $\mathcal{W}$-valued random variable with conditional probability mass function $\mathbb{P}[w_t|x, q]$ that we assume to satisfy the following:

(A1) $\mathbb{P}[w_{1:T}|x, q] = \prod_{i=1}^{T} \mathbb{P}[w_i|x, q]$,
(A2) $\mathbb{P}[w_i|x, q] = \mathbb{P}[y_i|x, q, v_i], \quad \forall i \in \{1, \ldots, T\}$,

where v_i is the chosen action and $w_{1:T}$ is a shorthand to denote the vector of sensor modes for the time horizon from $t = 1$ to $t = T$. Assumption (A1) represents the typical assumption that stochastic noise is conditionally independent over time, while Assumption (A2) means that conditioned on the state x, the fault mode q and the action v_i, there is a 1-to-1 mapping between y_i and w_i (typically true by the definition of w_i). These assumptions hold in the noisy Bayesian active learning problems in [19], [20].

Next, as in [11], [13], [21], we reduce the noisy problem to the noiseless case by considering "noisy" copies of the state, each corresponding to a fault mode and a noise mode for each t. More formally, we have *objects* consisting of a tuple of state, fault mode and noise modes, $\mathbf{x} \triangleq (x, q, w_{1:T}) \in$

$\mathcal{X} \times \mathcal{Q} \times \mathcal{W}^T \triangleq \boldsymbol{\mathcal{X}}$, and the state $x \in \mathcal{X}$ can be thought of as a *group* of objects, i.e., $x = \{(x, q, w_{1:T}) | q \in \mathcal{Q}, w_{1:T} \in \mathcal{W}^T\}$. The decision maker is allowed to take a sensing action (or run a test) $v \in \mathcal{V}$. The sensing action v generates a measurement observation/outcome in $y \in \mathcal{Y}$ whose value is (uniquely) determined by the true state x_o, fault mode q_o and noise mode $w_{t,o}$ at time t, i.e., $y_t = \mu(v, x_o, q_o, w_{t,o})$. Given a budget on the number of actions, T, the goal of the decision maker is to minimize the uncertainty in the estimation of the actual state x_o by an adaptive sequential choice of irrevocable actions based only on past (noisy) observations from previous actions. The goal of "minimum uncertainty" is described by a reward function $f : 2^{\mathcal{V}} \times \mathcal{X} \times \mathcal{Q} \times \mathcal{W}^T$ that is to be maximized, where we slightly abuse the notation of $2^{\mathcal{V}}$ to represent collections of sets that allow repeated elements.

To formalize the noisy discrete state estimation problem, we represent the pairs of actions $\{v_1, \ldots, v_t\}$ and observed outcomes $\{y_1, \ldots, y_t\}$ as the partial realization $\psi_t = \{(v_i, y_i)\}_{i \in \{1,\ldots,t\}}$. Given two partial realizations ψ_t and $\psi_{t'}$, we call ψ_t a subrealization of $\psi_{t'}$ if $\psi_t \subseteq \psi_{t'}$. Moreover, we define $D(y, v, q, w_t)$ as the set of states $x \in \mathcal{X}$ that gives the same observation y under the action v and (x, q, w_t) is the true tuple of state, fault mode and noise mode at time t. We then define $S_{t,q,w_{1:t}}$ as the set of all compatible states with the hypothesis that q and $w_{1:t}$ are the true fault and noise modes up to iteration t, i.e., the set of all states that produce the same set of outcomes $\{\mu(v_1, x_o, q_o, w_{1,o}), ..., \mu(v_t, x_o, q_o, w_{i,o})\}$ under the set of actions $\{v_1, ..., v_t\}$:

$$S_{t,q,w_{1:t}} = \cap_{i \in \{1,\ldots,t\}} D(\mu(v_i, x_o, q_o, w_{t,o}), v_i, q, w_i). \quad (1)$$

Since only intersections are taken, the order of actions v_i does not matter.

Furthermore, we define a policy π as a function of the observed partial realizations ψ_t to action v, i.e., $v_{t+1} = \pi(\psi_t)$. Randomized policies that specify the distribution on actions are also allowed.

The noisy discrete state estimation (henceforth referred to by its more common name, noisy Bayesian active diagnosis) problem is as follows:

Problem 1 (Noisy Bayesian Active Diagnosis)**.** *Given a reward function f and a probability mass function $\mathbb{P}[\mathbf{x}]$, the objective of the noisy Bayesian active diagnosis (or noisy discrete state estimation) problem is to find a policy π^* with a budget of T actions such that*

$$\begin{aligned} &\pi^* \in \arg\max_{\pi} \mathbb{E}[f(\tilde{\mathcal{V}}(\pi, \mathbf{X}), \mathbf{X})] \\ &\text{subject to} \quad |\tilde{\mathcal{V}}(\pi, \mathbf{x}_o)| \leq T, \forall \mathbf{x}_o \in \boldsymbol{\mathcal{X}}, \end{aligned} \quad (2)$$

with expectation taken with respect to (w.r.t.) $\mathbf{X}$ and $\tilde{\mathcal{V}}(\pi, \mathbf{x}_o) \subseteq 2^{\mathcal{V}}$ is the set of all action sequences under policy π with the state $\mathbf{x}_o$.

IV. Preliminaries

We begin by introducing some definitions for set functions. Note that we allow sets to have repeated elements and the union operation is defined accordingly.

Definition 1 (Conditional Expected Marginal Benefit [6])**.** *Given an objective function f, an action $v \in \mathcal{V}$ and a partial realization ψ_t, the* conditional expected marginal benefits *of an action v and a policy π conditioned on having observed ψ_t are defined as*

$$\begin{aligned} \Delta(v|\psi_t) &\triangleq \mathbb{E}[f(v_{1:t} \cup \{v\}, \mathbf{X}) - f(v_{1:t}, \mathbf{X})|\psi_t], \quad (3) \\ \Delta(\pi|\psi_t) &\triangleq \mathbb{E}[f(v_{1:t} \cup \tilde{\mathcal{V}}(\pi, \mathbf{x}_0), \mathbf{X}) - f(v_{1:t}, \mathbf{X})|\psi_t], \end{aligned}$$

respectively, with the expectation taken w.r.t. $\mathbb{P}[\mathbf{x}|\psi_t]$.

Definition 2 (Adaptive Monotonicity [6])**.** *A function $f : 2^{\mathcal{V}} \times \boldsymbol{\mathcal{X}} \to \mathbb{R}_{\geq 0}$ is* adaptive monotone *w.r.t. distribution $\mathbb{P}[\mathbf{x}]$ if for all $v \in \mathcal{V}$ and ψ_t with $\mathbb{P}[\psi_t] > 0$, $\Delta(v|\psi_t) \geq 0$.*

Definition 3 (ζ-Weak Adaptive Submodularity [13])**.** *A function $f : 2^{\mathcal{V}} \times \boldsymbol{\mathcal{X}} \to \mathbb{R}_{\geq 0}$ is ζ-*weakly adaptive submodular *w.r.t. distribution $\mathbb{P}[\mathbf{x}]$ if for all $\psi_t, \psi_{t'}$ such that ψ_t is a subrealization of $\psi_{t'}$, i.e., $\psi_t \subseteq \psi_{t'}$, and for all $v \in \mathcal{V} \setminus v_{1:t'}$,*

$$\Delta(v|\psi_{t'}) \leq \zeta \Delta(v|\psi_t),$$

for some adaptive submodularity factor $\zeta \geq 1$.

Adaptive monotonicity has the interpretation that the conditional expected marginal benefit of any action v is non-negative, while weak adaptive submodularity has the interpretation that the conditional expected marginal benefit of any fixed action (or query) v does not increase "too much" as more actions are performed and their measurements are observed. Note that the notion of ζ-weak adaptive submodularity generalizes the (strong) adaptive submodularity defined in [6], where $\zeta = 1$.

When the marginal gain $\Delta(v|\psi_{t'})$ satisfies adaptive monotonicity and weak adaptive submodularity, [13] recently showed that the following near-optimal performance guarantee can be obtained with an adaptive greedy algorithm.

Theorem 1 (Near-Optimality Guarantee [13])**.** *Fix any $\zeta \geq 1$. Let the greedy policy π_ℓ^{greedy} be run for ℓ iterations (i.e., it selects ℓ actions), and π_T^* be any policy selecting at most T actions for any realization $\mathbf{x}$. Then, for adaptive monotone and ζ-weakly adaptive submodular f (with $f(\emptyset) = 0$),*

$$f_{avg}(\pi_\ell^{greedy}) > (1 - e^{-\ell/\zeta T}) f_{avg}(\pi_T^*),$$

where $f_{avg}(\pi) \triangleq \mathbb{E}[f(\tilde{\mathcal{V}}(\pi, \mathbf{X}), \mathbf{X})]$ is the expected reward of the policy π w.r.t. $\mathbb{P}[\mathbf{x}]$.

V. Noisy Bayesian Active Diagnosis

The noisy Bayesian active diagnosis is inherently combinatorial and hence, computationally intractable for large problems. To circumvent this difficulty, researchers have often resorted to approximation algorithms to obtain suboptimal solutions reasonably fast but still with provable performance. For the Bayesian active diagnosis and learning problems, greedy algorithms are widely used when the reward function is (strongly) adaptive submodular [6] because of their near-optimal performance. However, many proposed reward functions for the noisy Bayesian active diagnosis and learning problems are, in general, not adaptive submodular.

TABLE I: Overview of reward functions for noisy Bayesian active diagnosis and their corresponding adaptive greedy policies and adaptive submodularity factors.

	REWARD FUNCTION, f_i	ADAPTIVE GREEDY POLICY, g_i ($v_{t+1} \in \arg\min_{v\in\mathcal{V}} g_i$)[a]	ADAP. SUBMODULARITY FACTOR[b]
(I)	$1 - \sum_{x\in\bigcup_{w_{1:t}\in\mathcal{W}^t}\bigcup_{q\in\mathcal{Q}} S_{t,q,w_{1:t}}} \mathbb{P}[x]$	$\sum_{y\in\mathcal{Y}} \tilde{g}(\cdot) \sum_{x\in\bigcup_{w_{1:t+1}\in\mathcal{W}^{t+1}}\bigcup_{q\in\mathcal{Q}} S_{t,q,w_{1:t}}\cap D(y,v,q,w_{t+1})} \mathbb{P}[x]$	$\zeta \leq \lvert\mathcal{Q}\rvert\lvert\mathcal{W}\rvert^T$
(II)	$1 - \sum_{x\in\bigcup_{w_{1:t}\in\mathcal{W}^t}\bigcup_{q\in\mathcal{Q}} S_{t,q,w_{1:t}}} \mathbb{P}[x]\mathbb{P}[y_{1:t}\lvert x, v_{1:t}]$	$\sum_{y\in\mathcal{Y}} \tilde{g}(\cdot) \sum_{x\in\bigcup_{w_{1:t+1}\in\mathcal{W}^{t+1}}\bigcup_{q\in\mathcal{Q}} S_{t,q,w_{1:t}}\cap D(y,v,q,w_{t+1})} \mathbb{P}[x]\mathbb{P}[y_{1:t}\lvert x, v_{1:t}]$	$\zeta \leq \lvert\mathcal{Q}\rvert$
(III)	$1-\sum_{q\in\mathcal{Q}} \sum_{x\in\bigcup_{w_{1:t}\in\mathcal{W}^t} S_{t,q,w_{1:t}}} \mathbb{P}[x,q]\mathbb{P}[y_{1:t}\lvert x,q,v_{1:t}]$	$\sum_{y\in\mathcal{Y}} (\tilde{g}(y,v,\psi_t,\{S_{t,q,w_{1:t}}\}))^2$	$\zeta = 1$

[a] Definition: $\tilde{g}(\cdot) = \tilde{g}(y,v,\psi_t,\{S_{t,q,w_{1:t}}\}) \triangleq \sum_{q\in\mathcal{Q}} \sum_{x\in\cup_{w_{1:t+1}\in\mathcal{W}^{t+1}} S_{t,q,w_{1:t}}\cap D(y,v,q,w_{t+1})} \mathbb{P}[x,q]\mathbb{P}[y_{1:t}\lvert x,q,v_{1:t}]$.

[b] These upper bounds can be derived analytically for the case when the fault and noise are uniformly distributed as described in Theorems 2 and 3, indicating that ζ is, in the worst case, linearly dependent on the number of faults and noise modes. Similar level of performance degradation due to ζ (cf. Theorem 1) was found in [11], [21]. For other fault and noise distributions, the upper bounds can be algorithmically computed using slightly modified versions of Algorithm 1 in [13].

Thus, proxy reward functions and corresponding algorithms were introduced in [4], [11], [21] to indirectly obtain performance guarantees for the original reward function, while [20] introduced a heuristic based on Extrinsic Jensen-Shannon (EJS) divergence and also provided performance guarantees when using that heuristic.

Recently, [13] introduced a simple yet valuable generalization of adaptive submodularity termed as *weakly adaptive submodularity*, for which a greedy algorithm with such reward functions also have near-optimal performance. More importantly, "sophisticated" proxy reward functions and algorithms (as in [11], [21]) are unnecessary, yet their performance guarantees are comparable. However, a reward function with this property was only proposed for the persistent fault setting without stochastic noise.

A. Overview

In this section, we extend the approach in [13], which is only applicable for persistent faults, to allow both persistent faults and non-persistent noise. When there are only persistent faults, our results reduce to those in [13], whereas when there is only stochastic noise, our results present submodularity-based alternatives to the noisy GBS and the EJS divergence heuristic in [19], [20]. This extension to consider both persistent faults and non-persistent noise is non-trivial because, depending on the type of noise, we may have to trade-off between the cost and benefit of repeating an action and exploring a new action at each step.

The reward functions we propose and their adaptive greedy policies and adaptive submodularity factors are summarized in Table I, and described below with proofs in the appendix.

1) Reward Functions: For the (persistently and non-persistently) noisy Bayesian active diagnosis problem, we consider three reward functions, $f_i(v_{1:T}, y_{1:T}, x, q, w_{1:t}), i \in \{\text{I}, \text{II}, \text{III}\}$:

$$\text{(i)}\quad f_{\text{I}}(\cdot) = 1 - \sum_{x\in\cup_{w_{1:t}\in\mathcal{W}^t}\cup_{q\in\mathcal{Q}} S_{t,q,w_{1:t}}} \mathbb{P}[x], \tag{4}$$

$$\text{(ii)}\quad f_{\text{II}}(\cdot) = 1 - \sum_{x\in\cup_{w_{1:t}\in\mathcal{W}^t}\cup_{q\in\mathcal{Q}} S_{t,q,w_{1:t}}} \mathbb{P}[x]\mathbb{P}[y_{1:t}\lvert x, v_{1:t}], \tag{5}$$

$$\text{(iii)}\quad f_{\text{III}}(\cdot) = 1 - \sum_{q\in\mathcal{Q}} \sum_{x\in\cup_{w_{1:t}\in\mathcal{W}^t} S_{t,q,w_{1:t}}} \mathbb{P}[x,q]\mathbb{P}[y_{1:t}\lvert x,q,v_{1:t}], \tag{6}$$

with $P[y_{1:0}|x,\emptyset] = 1$. The reward functions represent several objectives for noisy Bayesian active diagnosis that are relevant to robust control/decision-making problems in safety-critical and health-related applications, where worst-case scenarios need to be considered [22]. In such applications, the objective is to minimize uncertainty by eliminating as many states or hypotheses (e.g., of diseases) as possible that are incompatible with measured/observed outcomes. By contrast, the minimum MAP error and minimum entropy solutions (that are less uncertain in terms of probability but with potentially more states) would be less useful for finding the minimal set of compatible states. Moreover, the reward functions (4)–(6) are generalizations of the well-received reward function for version space mass reduction [6]; however, they are in general no longer (strongly) adaptive submodular.

Each proposed reward function has a different interpretation, based on whether the elimination criterion is based on the prior probability of the states x (Case I), based on the prior probability of the states x weighted by the likelihood for the observations (Case II) or based on the prior probability of the state-fault mode pair (x,q) weighted by the likelihood for the observations (Case III). They serve different purposes and we do not advocate for the significance of either one over the others.

2) An Adaptive Greedy Policy: Since we will show that the reward functions we propose are weakly adaptive submodular (a diminishing returns property), an adaptive greedy policy will provide near-optimal performance guarantees (cf. Theorem 1), while having the advantage of a polynomial-time computational complexity. Therefore, we consider adaptive greedy policies that, at each step, myopically and greedily maximize the expected gain based only on past (noisy) observations from previous actions:

$$\pi_i^{greedy}(\psi_t) \triangleq v_{t+1} \in \arg\max_{v\in\mathcal{V}} \Delta_i(v|\psi_t), \quad i \in \{\text{I}, \text{II}, \text{III}\},$$

where $\Delta_i(v|\psi_t)$ is derived according to Definition 1 based on the reward function f_i defined above. It can be easily

shown that this is equivalent to the following minimization of greedy loss functions g_i:

$$v_{t+1} \in \arg\min_{v\in\mathcal{V}} g_i(v, \psi_t, \{S_{t,q,w_{1:t}}\}), \quad i \in \{\text{I}, \text{II}, \text{III}\},$$

at every iteration/step t, where $g_i(\cdot)$ can be found in Table I (will be derived in Lemmas 4, 5 and 6 in Appendix B).

Interestingly, it can be verified that the worst-case complexity for all three adaptive greedy policies are the same (its proof is omitted due to space limitations).

Proposition 1 (Worst-Case Complexity)**.** *For all three adaptive greedy policies corresponding to reward functions f_{I}, f_{II} and f_{III}, the number of set operations per step t is $O(|\mathcal{Y}||\mathcal{V}||\mathcal{Q}||\mathcal{W}|)$, while the number of arithmetic operations per step t is $O(|\mathcal{Y}||\mathcal{V}||\mathcal{Q}||\mathcal{X}|)$.*

B. Theoretical Analysis

We now state our main results on the properties of the reward functions f_{I}, f_{II} and f_{III} in (4), (5), (6) and their corresponding near-optimal performance guarantees for the persistently and non-persistently noisy Bayesian active diagnosis problem (proofs will be provided in Appendix C).

1) Reward Function f_{I}:

Proposition 2 (Adaptive Monotonicity of f_{I})**.** *The reward function f_{I} in* (4) *is adaptive monotone.*

Proposition 3 (ζ_{I}-Weak Adaptive Submodularity of f_{I})**.** *The reward function f_{I} in* (4) *is ζ_{I}-weakly adaptive submodular with*

$$1 \le \zeta_{\text{I}} \le \overline{\zeta}_{\text{I}} \le \frac{|\mathcal{Q}||\mathcal{W}|^T}{\min\limits_{\{x\in\mathcal{X}, q\in\mathcal{Q}, w_{1:T}\in\mathcal{W}^T : \mathbb{P}[x,q,w_{1:T}]>0\}} \mathbb{P}[x,q,w_{1:T}]}, \quad (7)$$

where ζ_{I} and $\overline{\zeta}_{\text{I}}$ are defined as

$$\zeta_{\text{I}} \triangleq \max_{\substack{v_{1:T}\in\mathcal{V}^T, v\in\mathcal{V},\\ y\in\mathcal{Y}, t=1,\ldots,T}} \frac{\sum\limits_{x\in\bigcup\limits_{q\in\mathcal{Q}}\bigcup\limits_{w_{1:t+1}\in\mathcal{W}^{t+1}} \sum\limits_{S_{t,q,w_{1:t}}\cap D(y,v,q,w_{t+1})} \mathbb{P}[x]}{\sum\limits_{w_{1:t+1}\in\mathcal{W}^{t+1}}\sum\limits_{q\in\mathcal{Q}}\sum\limits_{x\in\sum\limits_{S_{t,q,w_{1:t}}\cap D(y,v,q,w_{t+1})}} \mathbb{P}[x,q,w_{1:t+1}]},$$

$$\overline{\zeta}_{\text{I}} \triangleq \max_{\substack{v_{1:T}\in\mathcal{V}^T, v\in\mathcal{V},\\ y\in\mathcal{Y}, t=1,\ldots,T}} \frac{\sum\limits_{w_{1:t+1}\in\mathcal{W}^{t+1}}\sum\limits_{q\in\mathcal{Q}}\sum\limits_{x\in\sum\limits_{S_{t,q,w_{1:t}}\cap D(y,v,q,w_{t+1})}} \mathbb{P}[x]}{\sum\limits_{w_{1:t+1}\in\mathcal{W}^{t+1}}\sum\limits_{q\in\mathcal{Q}}\sum\limits_{x\in\sum\limits_{S_{t,q,w_{1:t}}\cap D(y,v,q,w_{t+1})}} \mathbb{P}[x,q,w_{1:t+1}]}.$$

Moreover, when there are no informative priors on the persistent faults and stochastic noise, i.e., $\mathbb{P}[x,q,w_{1:t+1}] = \mathbb{P}[x]\frac{1}{|\mathcal{Q}||\mathcal{W}|^{t+1}}$, we have $\zeta_{\text{I}} \le \overline{\zeta}_{\text{I}} = |\mathcal{Q}||\mathcal{W}|^T$.

Theorem 2. *For any true state $x_o \in \mathcal{X}$, fault mode $q_o \in \mathcal{Q}$ and noise modes $w_{1:T,o} \in \mathcal{W}^T$, the adaptive greedy policy $\pi^{greedy}_{\text{I},T}$ for the reward function f_{I} in* (4) *guarantees that*

$$f_{\text{I},avg}(\pi^{greedy}_{\text{I},T}) > (1 - e^{-1/\zeta_{\text{I}}}) f_{\text{I},avg}(\pi^{*}_{\text{I},T}),$$

where $f_{\text{I},avg}(\pi^{*}_{\text{I},T})$ is achieved in T steps by the optimal policy and ζ_{I} is given in (7)*.*

Furthermore, without informative priors on the noise (i.e., $\zeta_{\text{I}} \le |\mathcal{Q}||\mathcal{W}|^T$), the adaptive greedy policy that selects T actions obtains at least $(1-e^{-1/|\mathcal{Q}||\mathcal{W}|^T})$ of the value of the optimal strategy that selects T actions.

2) Reward Function f_{II}:

Proposition 4 (Adaptive Monotonicity of f_{II})**.** *The reward function f_{II} in* (5) *is adaptive monotone.*

Proposition 5 (ζ_{II}-Weak Adaptive Submodularity of f_{II})**.** *The reward function f_{II} in* (4) *is ζ_{II}-weakly adaptive submodular with*

$$1 \le \zeta_{\text{II}} \le \overline{\zeta}_{\text{II}} \le \frac{|\mathcal{Q}|}{\min\limits_{\{x\in\mathcal{X}, q\in\mathcal{Q}, w_{1:T}\in\mathcal{W}^T : \mathbb{P}[x,q,w_{1:T}]>0\}} \mathbb{P}[x,q,w_{1:T}]}, \quad (8)$$

where ζ_{II} and $\overline{\zeta}_{\text{II}}$ are defined as

$$\zeta_{\text{II}} \triangleq \max_{\substack{v_{1:T}\in\mathcal{V}^T, v\in\mathcal{V},\\ y\in\mathcal{Y}, t=1,\ldots,T}} \frac{\sum\limits_{w_{1:t+1}\in\mathcal{W}^{t+1}}\sum\limits_{x\in\bigcup\limits_{q\in\mathcal{Q}}\sum\limits_{S_{t,q,w_{1:t}}\cap D(y,v,q,w_{t+1})} \mathbb{P}[x,w_{1:t+1}]}{\sum\limits_{w_{1:t+1}\in\mathcal{W}^{t+1}}\sum\limits_{q\in\mathcal{Q}}\sum\limits_{x\in\sum\limits_{S_{t,q,w_{1:t}}\cap D(y,v,q,w_{t+1})}} \mathbb{P}[x,q,w_{1:t+1}]},$$

$$\overline{\zeta}_{\text{II}} \triangleq \max_{\substack{v_{1:T}\in\mathcal{V}^T, v\in\mathcal{V},\\ y\in\mathcal{Y}, t=1,\ldots,T}} \frac{\sum\limits_{w_{1:t+1}\in\mathcal{W}^{t+1}}\sum\limits_{q\in\mathcal{Q}}\sum\limits_{x\in\sum\limits_{S_{t,q,w_{1:t}}\cap D(y,v,q,w_{t+1})}} \mathbb{P}[x,w_{1:t+1}]}{\sum\limits_{w_{1:t+1}\in\mathcal{W}^{t+1}}\sum\limits_{q\in\mathcal{Q}}\sum\limits_{x\in\sum\limits_{S_{t,q,w_{1:t}}\cap D(y,v,q,w_{t+1})}} \mathbb{P}[x,q,w_{1:t+1}]}.$$

Moreover, without informative priors on the persistent faults, i.e., $\mathbb{P}[x,q,w_{1:t+1}] = \mathbb{P}[x,w_{1:t+1}]\frac{1}{|\mathcal{Q}|}$, we have $\zeta_{\text{II}} \le \overline{\zeta}_{\text{II}} = |\mathcal{Q}|$.

Theorem 3. *For any true state $x_o \in \mathcal{X}$, fault mode $q_o \in \mathcal{Q}$ and noise modes $w_{1:T,o} \in \mathcal{W}^T$, the adaptive greedy policy $\pi^{greedy}_{\text{II},T}$ for the reward function f_{II} in* (5) *guarantees that*

$$f_{\text{II},avg}(\pi^{greedy}_{\text{II},T}) > (1 - e^{-1/\zeta_{\text{II}}}) f_{\text{II},avg}(\pi^{*}_{\text{II},T}),$$

where $f_{\text{II},avg}(\pi^{*}_{\text{II},T})$ is achieved in T steps by the optimal policy and ζ_{II} is given in (8)*.*

Furthermore, without informative priors on the persistent faults (i.e., when $\zeta_{\text{II}} \le |\mathcal{Q}|$), the adaptive greedy policy that selects T actions obtains at least $(1-e^{-1/|\mathcal{Q}|})$ of the value of the optimal strategy.

3) Reward Function f_{III}:

Proposition 6 (Adaptive Monotonicity of f_{III})**.** *The reward function f_{III} in* (6) *is adaptive monotone.*

Proposition 7 (Adaptive Submodularity of f_{III})**.** *The reward function f_{III} in* (6) *is adaptive submodular (i.e., 1-weakly adaptive submodular).*

Theorem 4. *For any true state $x_o \in \mathcal{X}$, fault mode $q_o \in \mathcal{Q}$ and noise modes $w_{1:T,o} \in \mathcal{W}^T$, the adaptive greedy policy $\pi^{greedy}_{\text{III},T}$ for the reward function f_{III} in* (6) *guarantees that*

$$f_{\text{III},avg}(\pi^{greedy}_{\text{III},T}) > (1 - e^{-1}) f_{\text{III},avg}(\pi^{*}_{\text{III},T}),$$

with $f_{\text{III},avg}(\pi^{*}_{\text{III},T})$ achieved in T steps by the optimal policy.

Remark 1. *The upper bounds, ζ_i, for any general fault and*

noise distributions can be algorithmically computed using Algorithm 1 in [13] with slight modifications. Moreover, the obtained upper bounds for the uniform case are comparable to the performance guarantees found in [11], [21] for (persistently) noisy active learning.

4) Special Cases: The results for the special cases below follow directly from the above analysis.

Stochastic Noise Only: When there is only stochastic noise, the following corollaries provide alternative submodularity-based adaptive greedy policies to the ones in [19], [20].

Corollary 1. *For any true state $x_o \in \mathcal{X}$, fault mode $q_o \in \mathcal{Q}$ and noise modes $w_{1:T,o} \in \mathcal{W}^T$, the adaptive greedy policy $\pi^{greedy}_{\text{IV},T}$ with the reward function $f_{\text{IV}}(v_{1:T}, y_{1:T}, x, w_{1:t}) = 1 - \sum_{x \in \cup_{w_{1:t} \in \mathcal{W}^t} S_{t,w_{1:t}}} \mathbb{P}[x]$ guarantees that*

$$f_{\text{IV},avg}(\pi^{greedy}_{\text{IV},T}) > (1 - e^{-1/\zeta_{\text{IV}}}) f_{\text{IV},avg}(\pi^{*}_{\text{IV},T}),$$

where $f_{\text{IV},avg}(\pi^{*}_{\text{IV},T})$ is achieved in T steps by the optimal policy and with ζ_{IV} that satisfies:

$$1 \leq \zeta_{\text{IV}} \leq \overline{\zeta}_{\text{IV}} \leq \frac{|\mathcal{W}|^T}{\min_{\{x \in \mathcal{X}, , w_{1:T} \in \mathcal{W}^T : \mathbb{P}[x, w_{1:T}] > 0\}} \mathbb{P}[x, w_{1:T}]},$$

where ζ_{IV} and $\overline{\zeta}_{\text{IV}}$ are defined as

$$\zeta_{\text{I}} \triangleq \max_{\substack{v_{1:T} \in \mathcal{V}^T, v \in \mathcal{V}, \\ y \in \mathcal{Y}, t=1,\ldots,T}} \frac{\sum_{x \in \bigcup_{w_{1:t+1} \in \mathcal{W}^{t+1}} S_{t,w_{1:t}} \cap D(y,v,w_{t+1})} \mathbb{P}[x]}{\sum_{w_{1:t+1} \in \mathcal{W}^{t+1}} \sum_{x \in S_{t,w_{1:t}} \cap D(y,v,w_{t+1})} \mathbb{P}[x, q, w_{1:t+1}]},$$

$$\overline{\zeta}_{\text{IV}} \triangleq \max_{\substack{v_{1:T} \in \mathcal{V}^T, v \in \mathcal{V}, \\ y \in \mathcal{Y}, t=1,\ldots,T}} \frac{\sum_{w_{1:t+1} \in \mathcal{W}^{t+1}} \sum_{x \in S_{t,w_{1:t}} \cap D(y,v,w_{t+1})} \mathbb{P}[x]}{\sum_{w_{1:t+1} \in \mathcal{W}^{t+1}} \sum_{x \in S_{t,w_{1:t}} \cap D(y,v,w_{t+1})} \mathbb{P}[x, w_{1:t+1}]}.$$

Moreover, with no informative priors on the stochastic noise (i.e., when $\zeta_{\text{IV}} \leq \overline{\zeta}_{\text{IV}} = |\mathcal{W}|^T$), the adaptive greedy policy that selects T actions obtains at least $(1 - e^{-1/|\mathcal{W}|^T})$ of the value of the optimal strategy that selects T actions.

Corollary 2. *For any true state $x_o \in \mathcal{X}$ and noise modes $w_{1:T,o} \in \mathcal{W}^T$, the adaptive greedy policy $\pi^{greedy}_{\text{V},T}$ with the reward function $f_{\text{V}}(v_{1:T}, y_{1:T}, x, w_{1:t}) = 1 - \sum_{x \in \cup_{w_{1:t} \in \mathcal{W}^t} S_{t,w_{1:t}}} \mathbb{P}[x]\mathbb{P}[y_{1:t}|x, v_{1:t}]$ guarantees that*

$$f_{\text{V},avg}(\pi^{greedy}_{\text{V},T}) > (1 - e^{-1}) f_{\text{V},avg}(\pi^{*}_{\text{V},T}),$$

with $f_{\text{V},avg}(\pi^{*}_{\text{V},T})$ achieved in T steps by the optimal policy.

Persistent Faults Only: The reward functions f_{I} and f_{II} with only persistent faults reduce to the one considered in [13] and have the same performance guarantees.

Noiseless Setting: The reward functions and performance guarantees reduce to the ones in [6].

VI. Illustrative Example

For an illustrative example, we return to the motivational example in Section II of an aircraft electrical system whose sensors are affected by persistent faults and stochastic noise.

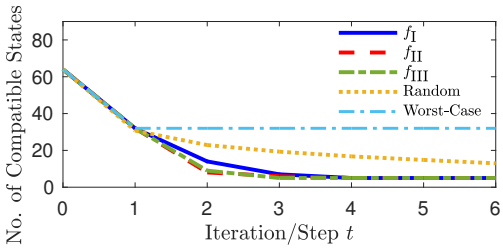

Fig. 2: Decreasing average number of compatible states with increasing number of iterations for each reward function: f_{I}, f_{II} and f_{III}, in comparison with uniformly random and worst-case strategies.

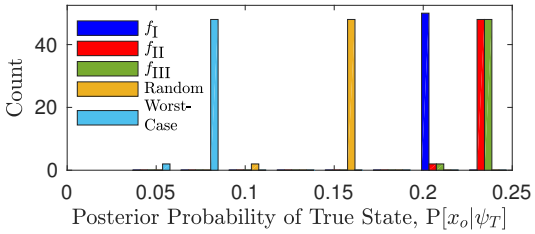

Fig. 3: Histogram of the final posterior probability of true state, $\mathbb{P}[x_o|\psi_T]$ where $T = 6$, over 50 runs for each reward function: f_{I}, f_{II} and f_{III}, in comparison with uniformly random and worst-case strategies.

As in [12], [13], our overall goal is to design a policy that sequentially/adaptively finds the set of compatible discrete states of the circuit by taking active sensing "actions" (i.e., opening or closing controllable contactors) and observing the sensor measurements. The availability of this minimal compatible set of discrete states will, for instance, allow the system operator to quickly narrow down which components need to be repaired.

By using four controllable contactors $\{C1, C3, C4, C6\}$ (thus, $|\mathcal{V}| = 2^4 = 16$ actions) and observing the readings of sensors $\{S1, S2, S3\}$, which may be noisy, the objective is to estimate the state of unknown components $\{G1, G2, R1, R2, C2, C5\}$ within a budget of $T = 6$ actions. For simplicity, we only allowed two states for the system components, i.e., *healthy* or *faulty*, and assumed that the state x was uniformly distributed on $\mathcal{X}$. Moreover, the sensor readings only took two values, i.e., *proper voltage* or *improper voltage*, where the observation of sensor 'S2' was prone to a persistent Type 1 fault with probability of obtaining the opposite outcome at $\mathbb{P}[q|x] = 0.2$ and the sensor 'S1' was corrupted by stochastic noise with probability of the opposite outcome, $\mathbb{P}[w_t|x, q] = \mathbb{P}[w_t] = 0.2$, for each step t.

To demonstrate the effectiveness of our approach using the three reward functions proposed in (4), (5) and (6), we compare their performances to a random strategy that takes actions uniformly at random and a worst-case strategy that repeatedly takes the *worst* action that maximizes (instead of minimizes) the number of compatible states. From our simulations, we observed that the adaptive greedy policy for all three reward functions that takes 6 actions (our budget) outperformed the uniformly random and worst-case strategies and more interestingly, performed equally well as

the exhaustive search policy that took all 16 actions (our benchmark). In fact, the number of compatible states was already equal to the minimum number of indistinguishable states with an exhaustive search after taking 4 actions. Hence, we compared the transient behavior of the number of compatible states at each iteration for each of the reward functions. Fig. 2 shows that the transient performances were comparable, with f_{I} being slightly tardier and with no clear winner between f_{II} and f_{III}.

Moreover, despite this not being the reward function we maximize, we compared the average posterior probability of the true state after taking 6 actions (over 50 runs) for each of the reward functions f_{I}, f_{II} and f_{III} with a fixed state x_o and fault mode q_o, but with different stochastic noise sampled from $\mathbb{P}[w_t]$ for each run. Somewhat surprisingly, the adaptive greedy policies for all 3 reward functions performed equally well as exhaustive searches. However, as we observe in Fig. 3, f_{I} yielded a posterior probability of 0.2 for all runs, while f_{II} and f_{III} had higher posterior probability during most runs. This is expected since only f_{II} and f_{III} take noise modes (the only quantity that was varied across runs) into account. In comparison, the uniformly random and worst-case strategies performed poorly, where the performance of the uniformly random strategy is approximately at the midpoint between the worst-case strategy and the adaptive greedy strategies with f_{II} and f_{III} (thus, also the exhaustive search policy).

VII. Conclusions

In this paper, we considered an extension of the noisy discrete state estimation problem (also known as noisy Bayesian active diagnosis) that allows the vector-valued observations to be simultaneously corrupted by both persistent sensor faults and non-persistent sensor noise. To this end, we introduced novel and meaningful reward functions for both faults and noise (as opposed to separate ones for different noise types). Moreover, we showed that these reward functions are both *adaptive monotone* and *weakly adaptive submodular*. Therefore, corresponding adaptive greedy policies were proposed to circumvent the complexity of the inherently combinatorial problem, which still have provable near-optimal performance guarantees. Our state estimation simulations for an aircraft electrical system with persistent sensor faults and stochastic sensor noise demonstrated that the adaptive greedy policy performs just as well as an exhaustive search policy while using significantly less computation.

References

[1] L.P. Kaelbling, M.L. Littman, and A.R. Cassandra. Planning and acting in partially observable stochastic domains. *Artificial intelligence*, 101(1):99–134, 1998.

[2] R.S. Sutton and A.G. Barto. *Reinforcement learning: An introduction*, volume 1. MIT Press Cambridge, 1998.

[3] A. Singh, A. Krause, C. Guestrin, W.J. Kaiser, and M.A. Batalin. Efficient planning of informative paths for multiple robots. In *IJCAI*, volume 7, pages 2204–2211, 2007.

[4] S. Javdani, Y. Chen, A. Karbasi, A. Krause, D. Bagnell, and S.S. Srinivasa. Near optimal bayesian active learning for decision making. In *AISTATS*, pages 430–438, 2014.

[5] N.I. Santoso, C. Darken, G. Povh, and J. Erdmann. Nuclear plant fault diagnosis using probabilistic reasoning. In *Power Engineering Society Summer Meeting, 1999. IEEE*, volume 2, pages 714–719. IEEE, 1999.

[6] D. Golovin and A. Krause. Adaptive submodularity: Theory and applications in active learning and stochastic optimization. *Journal of Artificial Intelligence Research*, 42:427–486, 2011.

[7] R. Debouk, S. Lafortune, and D. Teneketzis. On an optimization problem in sensor selection. *Discrete Event Dynamic Systems*, 12(4):417–445, 2002.

[8] P. Singh, S.Z. Yong, and E. Frazzoli. Supermodular batch state estimation in optimal sensor scheduling. *IEEE Control Systems Letters*, 2017.

[9] S. T. Jawaid and S. L. Smith. Submodularity and greedy algorithms in sensor scheduling for linear dynamical systems. *Automatica*, 61:282–288, 2015.

[10] T.S. Jaakkola and M.I. Jordan. Variational methods and the qmr-dt database. *NATO ASI Series F Computer and Systems Sciences*, 168:185–214, 1998.

[11] G. Bellala, S.K. Bhavnani, and C. Scott. Group-based active query selection for rapid diagnosis in time-critical situations. *IEEE Transactions on Information Theory*, 58(1):459–478, 2012.

[12] Q. Maillet, H. Xu, N. Ozay, and R.M. Murray. Dynamic state estimation in distributed aircraft electric control systems via adaptive submodularity. In *IEEE Conference on Decision and Control*, pages 5497–5503, 2013.

[13] S.Z. Yong, L. Gao, and N. Ozay. Weak adaptive submodularity and group-based active diagnosis with applications to state estimation with persistent sensor faults. In *IEEE American Control Conference (ACC)*, pages 2574–2581, May 2017.

[14] T.H. Summers and J. Lygeros. Optimal sensor and actuator placement in complex dynamical networks. *IFAC Proceedings Volumes*, 47(3):3784–3789, 2014.

[15] V. Tzoumas, A. Jadbabaie, and G.J. Pappas. Sensor placement for optimal Kalman filtering: Fundamental limits, submodularity, and algorithms. In *American Control Conference*, pages 191–196, 2016.

[16] U. Feige. A threshold of ln n for approximating set cover. *Journal of the ACM (JACM)*, 45(4):634–652, 1998.

[17] G. Bellala, J. Stanley, S.K. Bhavnani, and C. Scott. A rank-based approach to active diagnosis. *IEEE transactions on pattern analysis and machine intelligence*, 35(9):2078–2090, 2013.

[18] A.X. Zheng, I. Rish, and A. Beygelzimer. Efficient test selection in active diagnosis via entropy approximation. *arXiv preprint arXiv:1207.1418*, 2012.

[19] R. Nowak. Noisy generalized binary search. In *Advances in neural information processing systems*, pages 1366–1374, 2009.

[20] M. Naghshvar, T. Javidi, and K. Chaudhuri. Bayesian active learning with non-persistent noise. *IEEE Transactions on Information Theory*, 61(7):4080–4098, 2015.

[21] D. Golovin, A. Krause, and D. Ray. Near-optimal bayesian active learning with noisy observations. In *Advances in Neural Information Processing Systems*, pages 766–774, 2010.

[22] K. Zhou and J.C. Doyle. *Essentials of robust control*, volume 104. Prentice hall Upper Saddle River, NJ, 1998.

Appendix

A. Connection to Group-Based Active Diagnosis

We first show that the proposed reward functions are related to the reward function for *group-based active diagnosis* in (5) of [13] under Assumptions (A1) and (A2). In particular, f_{I} corresponds to the case when the group is the state x, f_{II} to the case with the tuple $(x, w_{1:T})$ as the group and f_{III} to the case with the tuple $(x, q, w_{1:T})$ as both the group and the object.

Lemma 1 (Reward function f_{I})**.** *Under Assumptions (A1) and (A2), with groups defined as states x and objects as tuples $(x, q, w_{1:T})$, the group-based reward function in (5) of [13] is equivalent to $f_{\text{I}} = 1 - \tilde{f}_{\text{I}}$, with*

$$\tilde{f}_{\text{I}}(v_{1:T}, y_{1:T}, x, q, w_{1:T}) \triangleq \sum_{x \in \cup_{w_{1:T} \in \mathcal{W}^T} \cup_{q \in \mathcal{Q}} S_{t,q,,w_{1:T}}} \mathbb{P}[x] = \sum_{x \in \cup_{w_{1:t} \in \mathcal{W}^t} \cup_{q \in \mathcal{Q}} S_{t,q,,w_{1:t}}} \mathbb{P}[x].$$

Proof. This follows since the set of compatible states is independent of future noise (Assumption (A1)). ∎

Lemma 2 (Reward function f_{II})**.** *Under Assumptions (A1) and (A2), with groups defined as $(x, w_{1:T})$ and objects as tuples $(x, q, w_{1:T})$, the group-based reward function in (5) of [13] is equivalent to $f_{\text{II}} = 1 - \tilde{f}_{\text{II}}$, with*

$$\tilde{f}_{\text{II}}(v_{1:T}, y_{1:T}, x, q, w_{1:T}) \triangleq \sum_{(x,w_{1:T})\in\cup_{q\in\mathcal{Q}} S_{t,q}} \mathbb{P}[x, w_{1:T}] = \sum_{x\in\cup_{w_{1:t}\in\mathcal{W}^t}\cup_{q\in\mathcal{Q}} S_{t,q,w_{1:t}}} \mathbb{P}[x]\mathbb{P}[y_{1:t}|x, v_{1:T}].$$

Proof. Starting from the reward function in [13], we can simplify $\tilde{f}_{\text{II}}(v_{1:T}, y_{1:T}, x, q, w_{1:T})$ to be

$$\begin{aligned}
&\tilde{f}_{\text{II}}(v_{1:T}, y_{1:T}, x, q, w_{1:T}) \\
&= \sum_{(x,w_{1:T})\in\cup_{q\in\mathcal{Q}} S_{t,q,w_{1:T}}} \sum_{q\in\mathcal{Q}} \mathbb{P}[x, q]\mathbb{P}[w_{1:T}|x, q] \\
&= \sum_{x\in \bigcup_{w_{1:T}\in\mathcal{W}^T} \bigcup_{q\in\mathcal{Q}} S_{t,q,w_{1:T}}} \sum_{q\in\mathcal{Q}} \sum_{w_{1:T}\in\mathcal{W}_{t,q,x}} \begin{pmatrix} \mathbb{P}[x, q]\mathbb{P}[w_{1:t}|x, q] \\ \mathbb{P}[w_{t+1:T}|x, q] \end{pmatrix} \\
&= \sum_{x\in \bigcup_{w_{1:T}\in\mathcal{W}^T} \bigcup_{q\in\mathcal{Q}} S_{t,q,w_{1:T}}} \sum_{q\in\mathcal{Q}} \begin{pmatrix} \mathbb{P}[x, q]\mathbb{P}[y_{1:t}|x, q, v_{1:T}] \\ \sum_{w_{t+1:T}\in\mathcal{W}^{T-t}} \mathbb{P}[w_{t+1:T}|x, q] \end{pmatrix} \\
&= \sum_{x\in \bigcup_{w_{1:t}\in\mathcal{W}^t} \bigcup_{q\in\mathcal{Q}} S_{t,q,w_{1:t}}} \sum_{q\in\mathcal{Q}} \mathbb{P}[x, q]\mathbb{P}[y_{1:t}|x, q, v_{1:T}] \\
&= \textstyle\sum_{x\in\cup_{w_{1:t}\in\mathcal{W}^t}\cup_{q\in\mathcal{Q}} S_{t,q,w_{1:t}}} \mathbb{P}[x]\mathbb{P}[y_{1:t}|x, v_{1:T}],
\end{aligned}$$

where the second equality is obtained by defining $\mathcal{W}_{t,q,x}$ as the set of noise modes that are compatible with ψ_t if (x, q) were the true state and mode, and by applying Assumption (A1). The third equality follows from Assumptions (A1) and (A2) and the fact that given $y_{1:t}$, x and q, $w_{1:t}$ is uniquely determined. ∎

Lemma 3 (Reward function f_{III})**.** *Under Assumptions (A1) and (A2), with the groups and objects being $(x, q, w_{1:T})$, the group-based reward function in (5) of [13] is equivalent to $f_{\text{III}} = 1 - \tilde{f}_{\text{III}}$, with*

$$\tilde{f}_{\text{III}}(v_{1:T}, y_{1:T}, x, q, w_{1:T}) \triangleq \sum_{(x,q,w_{1:T})\in S_t} \mathbb{P}[x, q, w_{1:T}] = \sum_{q\in\mathcal{Q}} \sum_{x\in \bigcup_{w_{1:T}\in\mathcal{W}^T} S_{t,q,w_{1:T}}} \mathbb{P}[x, q]\mathbb{P}[y_{1:t}|x, q, v_{1:T}].$$

Proof. As with Lemma 2, under Assumptions (A1) and (A2), $\tilde{f}_{\text{III}}(v_{1:T}, y_{1:T}, x, q, w_{1:T})$ simplifies to:

$$\begin{aligned}
&\tilde{f}_{\text{III}}(v_{1:T}, y_{1:T}, x, q, w_{1:T}) \\
&= \sum_{q\in\mathcal{Q}} \sum_{x\in\cup_{w_{1:T}\in\mathcal{W}^T} S_{t,q,w_{1:T}}} \sum_{w_{1:T}\in\mathcal{W}_{t,q,x}} \begin{pmatrix} \mathbb{P}[x, q]\mathbb{P}[w_{1:t}|x, q] \\ \mathbb{P}[w_{t+1:T}|x, q] \end{pmatrix} \\
&= \sum_{q\in\mathcal{Q}} \sum_{x\in\cup_{w_{1:T}\in\mathcal{W}^T} S_{t,q,w_{1:T}}} \begin{pmatrix} \mathbb{P}[x, q]\mathbb{P}[y_{1:t}|x, q, v_{1:T}] \\ \sum_{w_{t+1:T}\in\mathcal{W}^{T-t}} \mathbb{P}[w_{t+1:T}|x, q] \end{pmatrix} \\
&= \textstyle\sum_{q\in\mathcal{Q}} \sum_{x\in \bigcup_{w_{1:T}\in\mathcal{W}^T} S_{t,q,w_{1:T}}} \mathbb{P}[x, q]\mathbb{P}[y_{1:t}|x, q, v_{1:T}].
\end{aligned}$$

∎

B. Greedy Loss Functions

Having established the connection between group-based reward function [13] with our proposed reward functions f_{I}, f_{II} and f_{III}, the corresponding greedy loss functions in Table I can be derived from (9) of [13]. Moreover, simplifications that result in causal (i.e., nonanticipative) and computationally efficient algorithms can be obtained under Assumptions (A1) and (A2), stated below without proof for the sake of brevity.

Lemma 4 (Reward function f_{I})**.** *Under Assumptions (A1) and (A2), with the groups being only the state x, the greedy loss function $g_{\text{I}}(v, \psi_t, \{S_{t,q,w_{1:T}}\})$ for the reward function f_{I} is:*

$$\begin{aligned}
&g_{\text{I}}(v, \psi_t, \{S_{t,q,w_{1:T}}\}) \\
&= \sum_{y\in\mathcal{Y}} \sum_{\tilde{x}\in \bigcup_{(\tilde{q},\tilde{w}_{1:T})\in\mathcal{Q}\times\mathcal{W}^T} (S_{t,\tilde{q},\tilde{w}_{1:T}} \cap D(y,v,\tilde{q},\tilde{w}_{1:T}))} \mathbb{P}[\tilde{x}] \sum_{(q,w_{1:T})\in\mathcal{Q}\times\mathcal{W}^T} \sum_{x\in (S_{t,q,w_{1:T}} \cap D(y,v,q,w_{1:T}))} \mathbb{P}[x, q] \\
&= \sum_{y\in\mathcal{Y}} \tilde{g}(y, v, \psi_t, \{S_{t,q,w_{1:t}}\}) \sum_{x\in \bigcup_{w_{1:t+1}\in\mathcal{W}^{t+1}} \bigcup_{q\in\mathcal{Q}} S_{t,q,w_{1:t}} \cap D(y,v,q,w_{t+1})} \mathbb{P}[x],
\end{aligned}$$

where $\tilde{g}(\cdot) \triangleq \sum_{q\in\mathcal{Q}} \sum_{x\in \bigcup_{w_{1:t+1}\in\mathcal{W}^{t+1}} S_{t,q,w_{1:t}} \cap D(y,v,q,w_{t+1})} \mathbb{P}[x, q]\mathbb{P}[y_{1:t}|x, q, v_{1:t}]$.

Lemma 5 (Reward function f_{II})**.** *Under Assumptions (A1) and (A2), with groups being tuples of state and noise modes $(x, w_{1:T})$, the greedy loss function $g_{\text{II}}(v, \psi_t, \{S_{t,q,w_{1:t}}\})$ for f_{II} is (cf. $\tilde{g}(\cdot)$ in Lemma 4):*

$$\begin{aligned}
&g_{\text{II}}(v, \psi_t, \{S_{t,q,w_{1:t}}\}) \\
&= \sum_{y\in\mathcal{Y}} \sum_{(\tilde{x},\tilde{w}_{1:T})\in \bigcup_{\tilde{q}\in\mathcal{Q}} (S_{t,\tilde{q}} \cap D(y,v,\tilde{q}))} \mathbb{P}[\tilde{x}, \tilde{w}_{1:T}] \sum_{q\in\mathcal{Q}} \sum_{(x,w_{1:T})\in (S_{t,q} \cap D(y,v,q))} \mathbb{P}[x, q, w_{1:T}] \\
&= \sum_{y\in\mathcal{Y}} \tilde{g}(y, v, \psi_t, \{S_{t,q,w_{1:t}}\}) \sum_{x\in \bigcup_{w_{1:t+1}\in\mathcal{W}^{t+1}} \bigcup_{q\in\mathcal{Q}} \left(S_{t,q,w_{1:t}} \cap D(y,v,q,w_{t+1})\right)} \mathbb{P}[x]\mathbb{P}[y_{1:t}|x, v_{1:t}].
\end{aligned}$$

Lemma 6 (Reward function f_{III})**.** *Under Assumptions (A1) and (A2), with the groups and objects being $(x, q, w_{1:T})$, the greedy loss function $g_{\text{III}}(v, \psi_t, \{S_{t,q,w_{1:t}}\})$ for f_{III} is (cf. $\tilde{g}(\cdot)$ in Lemma 4):*

$$\begin{aligned}
&g_{\text{III}}(v, \psi_t, \{S_{t,q,w_{1:t}}\}) \\
&= \sum_{y\in\mathcal{Y}} \sum_{(\tilde{x},\tilde{q},\tilde{w}_{1:T})\in S_t\cap D(y,v)} \mathbb{P}[\tilde{x}, \tilde{q}, \tilde{w}_{1:T}] \sum_{(x,q,w_{1:T})\in S_t\cap D(y,v)} \mathbb{P}[x, q, w_{1:T}] \\
&= \textstyle\sum_{y\in\mathcal{Y}} (\tilde{g}(y, v, \psi_t, \{S_{t,q,w_{1:t}}\}))^2.
\end{aligned}$$

C. Proofs of Propositions 2–7 and Theorems 2–4

By the equivalence of f_{I}, f_{II} and f_{III} with the group-based reward function in [13] (Lemmas 1, 2 and 3), it follows that Propositions 2 & 3, 4 & 5 and 6 & 7 and thus, Theorems 2, 3 and 4 hold from the results in [13].