Submodular Surrogates for Value of Information

Yuxin Chen
ETH Zürich
yuxin.chen@inf.ethz.ch

Shervin Javdani
Carnegie Mellon University
sjavdani@cmu.edu

Amin Karbasi
Yale University
amin.karbasi@yale.edu

J. Andrew Bagnell
Carnegie Mellon University
dbagnell@ri.cmu.edu

Siddhartha Srinivasa
Carnegie Mellon University
ss5@andrew.cmu.edu

Andreas Krause
ETH Zürich
krausea@ethz.ch

Abstract

How should we gather information to make effective decisions? A classical answer to this fundamental problem is given by the decision-theoretic value of information. Unfortunately, optimizing this objective is intractable, and myopic (greedy) approximations are known to perform poorly. In this paper, we introduce DIRECT, an efficient yet near-optimal algorithm for nonmyopically optimizing value of information. Crucially, DIRECT uses a novel surrogate objective that is (1) aligned with the value of information problem; (2) efficient to evaluate and (3) adaptive submodular. This latter property enables us to utilize efficient greedy optimization while providing strong approximation guarantees. We extensively demonstrate the utility of our approach on three diverse case-studies: active learning for interactive content search, optimizing value of information in conservation management, and touch-based robotic localization. On the latter application, we demonstrate DIRECT in closed-loop on an actual robotic platform.

1 Introduction

In many real-world decision making tasks we must adaptively choose among informative but expensive tests. As an illustrative example, consider medical diagnosis [1], where many medical tests are available, each with a different cost. It is important to administer tests that will enable us to provide the most effective treatment. In such systems, the reward of making a decision depends on some unknown hidden state (e.g., the patient's condition). Generally, it is impossible to observe this hidden state directly, but one can choose to perform tests – observe the outcome of variables correlated with the hidden state – but at some cost. The task is then to find a policy to select the most informative tests, so that we can gather enough information to make effective decisions, while minimizing the cost of testing. Similar problems arise in numerous other domains, ranging from optimal experimental design [2] to recommender systems [3] to policy making [4].

Related work A classical approach to information gathering for decision making is the decision-theoretic *value of information* [5]. Here, we seek policies that maximize the increase in the maximum expected utility that the decision maker could obtain when acting upon the acquired information. Optimizing this criterion in general probabilistic models is NP^{PP}-complete [6]. Consequently, greedy heuristics that myopically select the next test are employed. It is known [7] that these heuristics can perform arbitrarily poorly; unfortunately exact algorithms for *non-myopic* value of information have so far been restricted to simple probabilistic models [6].

The problem of selecting information gathering tests for purely *reducing uncertainty about some hidden variable* (ignoring utilities of decision making) is studied in the context of active learning [8, 9, 10, 11] and (Bayesian) experimental design [2]. Deriving optimal policies is generally NP-hard [12], but some approximation results are known. In particular, if tests are noise-free (i.e., deterministic functions of the hidden state), the problem is known as the Optimal Decision Tree (ODT) problem, and a simple greedy algorithm, called generalized binary search (GBS), is guaranteed to produce a bounded approximation to the optimal policy in terms of the cost [13].

Recently, these results have been brought closer to decision making by associating each hidden state with some optimal decision(s), and seeking policies that reduce the uncertainty about the hidden state only to the extent to make the right decision. Two algorithms, namely *equivalence class edge cutting* (EC^2) [14] and *hyperedge cutting* (HEC) [3] provide approximation guarantees for this problem. Since our approach builds on these techniques, we review them in more detail in Section 1.

Our contributions In this paper, we provide a principled framework for a class of *non-myopic value of information* problems: We seek a policy of minimal cost, which guarantees that upon termination, a near-optimal decision – one that provides almost as much utility as achievable by carrying out *all* tests – is identified. Instead of optimizing the classical decision-theoretic value of information, we propose DIRECT, an efficient surrogate objective function. We show that it exhibits *adaptive submodularity* [15], a natural diminishing returns property, generalizing the classical notion of submodularity to adaptive policies. This result allows us to greedily maximize the surrogate, while still maintaining a strong theoretical guarantee. Experimental results show that our algorithm significantly outperforms myopic value of information in most settings. Moreover, our algorithm is exponentially faster than HEC in theory, significantly faster (often by orders of magnitude) in practice, while offering similar empirical performance.

2 Background and Problem Statement

The Value of Information and Decision Region Determination Problem. Assume that there is some unknown hidden discrete random variable $Y \in \mathcal{Y}$ upon which we want to make a decision. In our medical diagnostics example, Y may represent the condition of the patient. We are given a set $\mathcal{T} = \{1, \ldots, n\}$ of possible (e.g., medical) tests; performing each test $t \in \mathcal{T}$ incurs a certain cost of $c(t) > 0$ and produces an outcome that is correlated with Y. We model the outcome of each test t by a discrete random variable $X_t \in \mathcal{X}$ and denote its observed outcome by x_t. Hereby, $\mathbf{x}_\mathcal{A} \in \mathcal{X}^\mathcal{A}$ is a vector of outcomes indexed by a set of tests $\mathcal{A} \subseteq \mathcal{T}$ that we have performed, and y is the realized value of the hidden variable Y. Further assume that there is a known prior distribution $\mathbb{P}[Y, X_1, \ldots, X_n]$ over the hidden variable and test outcomes admitting efficient inference, i.e., we can compute the posterior distribution $\mathbb{P}[Y = y \mid \mathbf{x}_\mathcal{A}]$ efficiently after having observed any $\mathbf{x}_\mathcal{A}$.

Suppose there is a finite set $\mathcal{D}$ of decisions to choose from. After performing a set of tests and observing their outcomes, we want to make the best decision given our belief about the hidden variable Y (e.g., we must decide how to treat the patient). Formally, we quantify the benefit of making a decision $d \in \mathcal{D}$ for any $y \in \mathcal{Y}$ by a utility function $u : \mathcal{Y} \times \mathcal{D} \to \mathbb{R}_{\geq 0}$. The expected value of a decision d after observing $\mathbf{x}_\mathcal{A}$ is $U(d \mid \mathbf{x}_\mathcal{A}) = \mathbb{E}_y[u(y, d) \mid \mathbf{x}_\mathcal{A}]$. The value of a specific set of observations $\mathbf{x}_\mathcal{A}$ is then defined as: $\mathrm{VoI}(\mathbf{x}_\mathcal{A}) = \max_{d \in \mathcal{D}} U(d \mid \mathbf{x}_\mathcal{A})$, i.e., the maximum expected utility achievable when acting upon observations $\mathbf{x}_\mathcal{A}$.

Consider performing *all* tests, receiving outcomes $\mathbf{x}_\mathcal{T}$, and making the most informed decision possible. This would achieve a value of $\mathrm{VoI}(\mathbf{x}_\mathcal{T})$. However, it may be possible to achieve nearly $\mathrm{VoI}(\mathbf{x}_\mathcal{T})$ with far fewer tests. Our goal is to adaptively select the cheapest tests to do so. Formally, we define the *regret* of a decision d given observations $\mathbf{x}_\mathcal{A}$ by $R(d \mid \mathbf{x}_\mathcal{A}) = \max_{\mathbf{x}_\mathcal{T}:\mathbb{P}[\mathbf{x}_\mathcal{T} \mid \mathbf{x}_\mathcal{A}]>0}[\mathrm{VoI}(\mathbf{x}_\mathcal{T}) - U(d \mid \mathbf{x}_\mathcal{T})]$. This regret bounds our loss in expected utility by stopping upon observing $\mathbf{x}_\mathcal{A}$ and committing to action d. Our goal is to find a policy π of minimum cost with regret of at most ε. Formally, a policy is a partial mapping from observation vectors $\mathbf{x}_\mathcal{A}$ to tests, specifying which test to run next (or that we should stop testing if $\mathbf{x}_\mathcal{A}$ is not in the domain of π) for any observation vector $\mathbf{x}_\mathcal{A}$. If variables $X_1, \ldots, X_n$ would result in outcomes $\mathbf{x}_\mathcal{T}$, we will obtain a set of observations, denoted as $\mathcal{S}(\pi, \mathbf{x}_\mathcal{T}) \subseteq \mathcal{T} \times \mathcal{X}$, by running policy π until termination (likely before exhausting all tests). The expected cost of a policy π is $\mathrm{cost}(\pi) = \mathbb{E}_{\mathbf{x}_\mathcal{T}}[c(\mathcal{S}(\pi, \mathbf{x}_\mathcal{T}))]$, where $c(\mathcal{S}(\pi, \mathbf{x}_\mathcal{T}))$ is the total cost of all tests run by π in the event $\mathbf{x}_\mathcal{T}$. Fix some small tolerance $\varepsilon \geq 0$. We seek a policy π^* with minimum cost, such that upon termination, π^* will suffer regret of at most ε:

$$\pi^* \in \arg\min_\pi \mathrm{cost}(\pi), \text{ s.t.} \forall \mathbf{x}_\mathcal{T} \; \exists d : R(d \mid \mathcal{S}(\pi, \mathbf{x}_\mathcal{T})) \leq \varepsilon \text{ whenever } \mathbb{P}[\mathbf{x}_\mathcal{T}] > 0. \quad (1)$$

In other words, we require that each feasible policy satisfies the following condition: Upon termination, we must be able to commit to a decision, such that we lose *at most* ε expected utility, compared to the optimal decision we could have made if we had also observed *all remaining* unobserved variables (assuming they are consistent with our observations). We call Problem (1) the *nonmyopic value of information problem for achieving near-maximal utility (NVOI-NMU).*

Importantly, this problem reduces[1] to a problem known as the *Decision Region Determination* (DRD) problem [3]. In DRD, we are given (1) a set of hypotheses $\mathcal{H} = \{h_1, \ldots, h_N\}$; (2) a random variable H distributed over $\mathcal{H}$ with known distribution $\mathbb{P}$; (3) a set of tests modeled as deterministic functions $f_1, \ldots f_n : \mathcal{H} \to \mathcal{X}$; (4) a cost function $c : \{1, \ldots, n\} \to \mathbb{R}_+$ and (5) a collection of subsets $\mathcal{R}_1, \ldots, \mathcal{R}_m \subseteq \mathcal{H}$ called *decision regions*. We seek a policy π^* of minimum cost, which adaptively picks tests i, observes their outcomes $X_i = f_i(H)$, where $H \in \mathcal{H}$ is the unknown hypothesis, such that upon termination, there exists at least one decision region that contains all hypotheses consistent with the observations made by the policy. That is, we seek

$$\pi^* \in \arg\min_{\pi} \mathrm{cost}(\pi), \text{s.t. } \forall h\ \exists d : \mathcal{H}(\mathcal{S}(\pi, h)) \subseteq \mathcal{R}_d. \tag{2}$$

Hereby $h \in \mathcal{H}$, and $\mathcal{H}(\mathbf{x}_\mathcal{A}) = \{h' \in \mathcal{H} : (i, x) \in \mathbf{x}_\mathcal{A} \Rightarrow f_i(h') = x\}$ is the set of hypotheses consistent with $\mathbf{x}_\mathcal{A}$. To reduce the NVOI-NMU Problem (1) to DRD (2), we interpret every outcome vector $\mathbf{x}_\mathcal{T}$ with positive probability as a hypothesis h. The interpretation of the prior, tests, and costs are immediate. It remains to define the decision regions. For each decision d, we set $\mathcal{R}_d$ to be the set of outcome vectors, for which d is an ε-optimal action, or formally, $\mathcal{R}_d = \{\mathbf{x}_\mathcal{T} : U(d \mid \mathbf{x}_\mathcal{T}) \geq \mathrm{VoI}(\mathbf{x}_\mathcal{T}) - \varepsilon\}$.

Existing approaches for solving the DRD problem. As a special case of the DRD problem, the *Equivalence Class Determination* (ECD) problem [14] only allows *disjoint* decision regions, i.e., $\mathcal{R}_i \cap \mathcal{R}_j = \emptyset$ for $i \neq j$. The EC^2 algorithm [14] considers hypotheses as nodes in a graph $G = (V, E)$, and defines weighted edges between hypotheses in different decision regions: $E = \cup_{i \neq j}\{\{h, h'\} : h \in \mathcal{R}_i, h' \in \mathcal{R}_j\}$, where the weight of an edge is defined as $w(\{h, h'\}) = \mathbb{P}[h] \cdot \mathbb{P}[h']$; similarly, the weight of a set of edges is $w(E') = \sum_{e \in E'} w(e)$. A test t with outcome x_t is said to cut edges $E(x_t) = \{\{h, h'\} \in E : f_t(h) \neq x_t \vee f_t(h') \neq x_t\}$. We aim to cut all edges that are incident to inconsistent hypotheses while minimizing the expected cost incurred.

The EC^2 objective is defined as the total weight of edges cut: $f_{EC}(\mathbf{x}_\mathcal{A}) := w\big(\cup_{t \in \mathcal{A}} E(x_t)\big)$. EC^2 is known to be near-optimal for the ECD problem. This result relies on the fact that f_{EC} is *adaptive submodular*, and *strongly adaptive monotone* [15]. Let $\mathbf{x}_\mathcal{A}$ and $\mathbf{x}_\mathcal{B}$ be two observation vectors. We call $\mathbf{x}_\mathcal{A}$ a *subrealization* of $\mathbf{x}_\mathcal{B}$, denoted as $\mathbf{x}_\mathcal{A} \preceq \mathbf{x}_\mathcal{B}$, if the index set $\mathcal{A} \subseteq \mathcal{B}$ and $\mathbb{P}[\mathbf{x}_\mathcal{B} \mid \mathbf{x}_\mathcal{A}] > 0$. A function $f : 2^{\mathcal{T} \times \mathcal{X}} \to \mathbb{R}$ is called *adaptive submodular* w.r.t. a distribution $\mathbb{P}$, if for any $\mathbf{x}_\mathcal{A} \preceq \mathbf{x}_\mathcal{B}$ and any test t it holds that $\Delta(t \mid \mathbf{x}_\mathcal{A}) \geq \Delta(t \mid \mathbf{x}_\mathcal{B})$, where $\Delta(t \mid \mathbf{x}_\mathcal{A}) := \mathbb{E}_{x_t}\big[f(\mathbf{x}_{\mathcal{A} \cup \{t\}}) - f(\mathbf{x}_\mathcal{A}) \mid \mathbf{x}_\mathcal{A}\big]$ (i.e., "adding information earlier helps more"). Further, function f is called *strongly adaptively monotone* w.r.t. $\mathbb{P}$, if for all $\mathcal{A}$, $t \notin \mathcal{A}$, and $x_t \in \mathcal{X}$, it holds that $f(\mathbf{x}_\mathcal{A}) \leq f(\mathbf{x}_{\mathcal{A} \cup \{t\}})$ (i.e., "adding information never hurts"). For decision problems satisfying adaptive submodularity and strongly adaptive monotonicity, the policy that greedily, upon having observed $\mathbf{x}_\mathcal{A}$, selects the test $t^* \in \arg\max_t \Delta(t \mid \mathbf{x}_\mathcal{A})/c(t)$, is guaranteed to attain near-minimal cost [15].

EC^2 crucially relies on the fact that decision regions are *disjoint*. In the presence of overlapping regions, there is no principled way to apply EC^2. Recently, the HEC algorithm [3] was proposed for solving the general DRD problem. It does so by creating an alternate representation – a hypergraph for splitting decision regions. The computational bottleneck for HEC lies in the construction of this hypergraph, where computation cost grows *exponentially* with the hyperedge cardinality, which depends on the maximum number of optimal decisions one can make for a hypothesis. Thus, when we have large overlap between regions – the common case for NVOI-NMU, in particular with larger ε – HEC becomes intractable.

3 The Decision Region Edge Cutting Algorithm

We now develop an *efficient yet near-optimal* criterion, namely *Decision Region Edge Cutting* (DIRECT), for solving the DRD – and hence the NVOI-NMU – problem.

The Noisy-OR Construction Suppose there are m possible decisions: $|\mathcal{D}| = m$. Our strategy will be to reduce the DRD problem to $O(m)$ instances of the ECD problem, such that solving

[1]The NVOI-NMU and DRD problems are in fact equivalent.

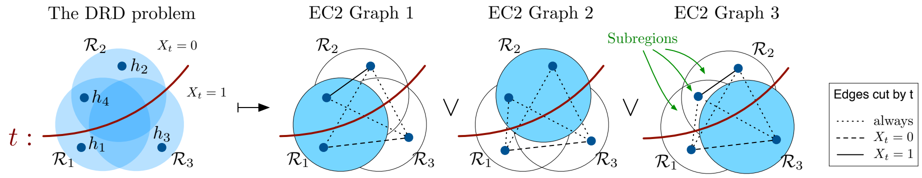


Figure 1: A toy DRD problem with three decision regions $\{\mathcal{R}_1, \mathcal{R}_2, \mathcal{R}_3\}$, and four possible hypotheses $\{h_1, h_2, h_3, h_4\}$. t is a test with two possible outcomes: $f_t(h_1) = f_t(h_3) = 1$ and $f_t(h_2) = f_t(h_4) = 0$. For each possible decision we can make, we construct a separate ECD problem: The three figures on the right illustrate the EC^2 graphs for each of the ECD problems. We can successfully make an optimal decision once one of the graphs is fully cut: e.g., if $X_t = 0$, graph 2 is fully cut, and we identify the optimal decision d_2.

any one of them is sufficient for solving the DRD problem. Concretely, we construct m different graphs, one for each decision. The role of graph i is to determine whether the unknown hypothesis h^* is contained in decision region $\mathcal{R}_i$ or not. Thus we aim to distinguish all the hypotheses in this decision region from the rest. To achieve this, we model graph i as an ECD problem, with one of the decision regions being $\mathcal{R}_i$. Further, we partition the remaining set of hypotheses $\mathcal{H} \setminus \mathcal{R}_i$ into a collection of *subregions*, such that within each subregion, all hypotheses are contained in exactly the same collection of decision regions from the original DRD problem. All the subregions are disjoint by definition, and hence we have a well-defined ECD problem. Solving this problem amounts to cutting all the edges between $\mathcal{R}_i$ and the subregions. See Figure 1 for illustration.

Notice that in this ECD problem, once all the edges are cut, either i is the optimal decision, or one of the subregions encodes the optimal decision. Therefore, optimizing the ECD problem associated with one of the m graphs is a *sufficient condition* for identifying the optimal decision.

Further notice that, among the m ECD problems associated with the m graphs, at least one of them has to be solved (i.e., all edges cut) before we uncover the optimal decision. Therefore, we get a *necessary condition* of the DRD constraints: we have to cut all the edges in *at least one* of the m graphs. This motives us to apply a logical OR operation on the m optimization problems. Denote the EC^2 objective function for graph i as f^i_{EC}, and normalize them so that $f^i_{EC}(\emptyset) = 0$ corresponds to observing nothing and $f^i_{EC}(\mathbf{x}_\mathcal{T}) = 1$ corresponds to all edges being cut. We combine the objective functions $f^1_{EC}, \dots, f^m_{EC}$ using a *Noisy-OR formulation*:

$$f_{DRD}(\mathbf{x}_\mathcal{A}) = 1 - \prod_i^m \left(1 - f^i_{EC}(\mathbf{x}_\mathcal{A})\right) \tag{3}$$

Note that by design $f_{DRD}(\mathbf{x}_\mathcal{A}) = 1$ iff $f^i_{EC}(\mathbf{x}_\mathcal{A}) = 1$ for *at least* one i. Thus, the DRD (and hence NVOI-NMU) Problem is formally equivalent to the following problem:

$$\pi^* \in \arg\min_\pi \mathrm{cost}(\pi), \text{ s.t.} \forall \mathbf{x}_\mathcal{T} : f_{DRD}(\mathcal{S}(\pi, \mathbf{x}_\mathcal{T})) \geq 1 \text{ whenever } \mathbb{P}[\mathbf{x}_\mathcal{T}] > 0. \tag{4}$$

The crucial advantage of this new formulation is given by the following Lemma:

Lemma 1. *f_{DRD} is is strongly adaptive monotone, and adaptive submodular w.r.t. $\mathbb{P}$.*

That is, the Noisy-OR formulation for multiple EC^2 functions preserves adaptive submodularity[2]. The proof of this result can be found in the supplemental material. These properties make f_{DRD} amenable for efficient greedy optimization. Formally, let $\Delta_{f_{DRD}}(t \mid \mathbf{x}_\mathcal{A}) := \mathbb{E}_{x_t}\left[f_{DRD}(\mathbf{x}_{\mathcal{A}\cup\{t\}}) - f_{DRD}(\mathbf{x}_\mathcal{A}) \mid \mathbf{x}_\mathcal{A}\right]$ be the expected marginal benefit in f_{DRD} by adding test t to $\mathbf{x}_\mathcal{A}$. With f_{DRD}, we can associate a greedy algorithm: It starts with the empty set, and at each iteration, having already observed $\mathbf{x}_\mathcal{A}$, selects the test t^* with the largest benefit-to-cost ratio: $t^* \in \arg\max_t \Delta_{f_{DRD}}(t \mid \mathbf{x}_\mathcal{A})/c(t)$. Since f_{DRD} is adaptive submodular, we can use lazy evaluation [15] to speed up the greedy selection process, while having the following guarantee:

Theorem 2. *Let m be the number of decisions, and π_{DRD} be the adaptive greedy policy w.r.t. the objective function Eq.* (3)*. Then it holds that* $\mathrm{cost}(\pi_{DRD}) \leq (2m \ln(1/p_{\min}) + 1)\,\mathrm{cost}(\pi^*)$*, where $p_{\min} = \min_{h\in\mathcal{H}} \mathbb{P}[h]$ is the minimum prior probability of any set of observations, and π^* is the optimal policy for Problem* (4)*, and hence also the NVOI-NMU and DRD Problems.*

[2]Similar constructions have been used for classical submodular set functions [16, 17], utilizing the fact that $f = 1 - \prod_i^m (1 - f_i)$ is submodular if each f_i is submodular. However, the function f is *not* necessarily adaptive submodular, even when each f_i is adaptive submodular and strongly adaptively monotone.

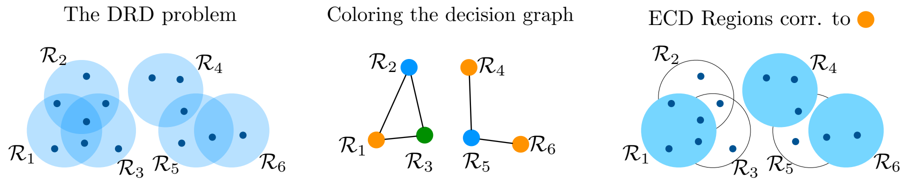


Figure 2: Reducing the cost upper bound via graph coloring. We only need to construct 3 ECD instances to compute f_{DRD}, instead of 6. The middle figure shows a possible coloring assignment on the decision graph of the DRD problem. On the right, we show one example ECD problem instancew with 7 disjoint (sub)regions.

This result follows from Lemma 1 and the general performance analysis of the greedy policy for adaptive submodular problems by [15]. The bound of the greedy algorithm is linear in the number of decision regions. Here the factor m is a result of taking the product of m EC^2 instances. In the following, we show how this bound can often be improved.

Improving the bound via Graph Coloring For certain applications, the number of decisions m can be large. Instead of constructing one ECD problem for each possible optimal decision separately, we can construct one ECD problem for several *non-overlapping* decision regions at once. Problem 4 remains to be equivalent to the DRD problem, as long as every decision region is accounted for by at least one of the ECD problems. See Figure 2 for illustration.

Formally, we construct an undirected graph $\mathcal{G} := \{\mathcal{D}, \mathcal{E}\}$ over all decision regions, where we establish an edge between any pair of overlapping decision regions. Finding a minimal set of non-overlapping decision region sets that covers all the decisions is equivalent to solving a graph coloring problem, where the goal is to color the vertices of the graph $\mathcal{G}$, such that no two adjacent vertices share the same color, using as few colors as possible. Thus, we can construct one ECD problem for all the decision regions of the same color, resulting in r different instances, and then use the Noisy-OR formulation to assemble these objective functions. That gives us the following theorem:

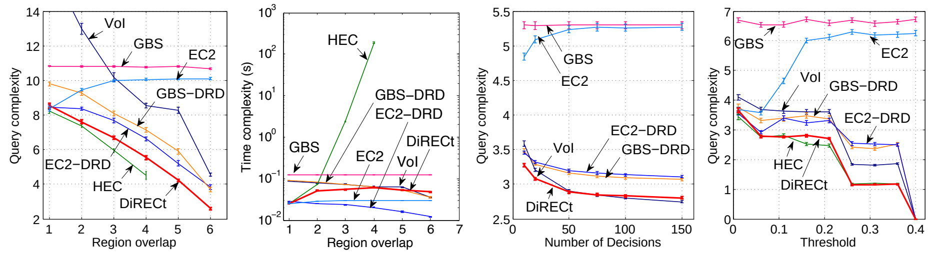
Theorem 3. *Let π_{DRD} be the adaptive greedy policy w.r.t. the objective function Eq.* (3)*, which is computed over ECD problem instances obtained via graph coloring. Let r be the number of colors used. Then it holds that $\mathrm{cost}(\pi_{DRD}) \leq (2r \ln(1/p_{\min}) + 1)\,\mathrm{cost}(\pi^*)$, where $p_{\min}$ is the minimum prior probability of any set of observations, and π^* is the optimal policy.*

While obtaining minimum graph colorings is NP-hard in general, one can show that every graph can be efficiently colored with at most one more color than the maximum vertex degree, denoted by deg, using a greedy coloring algorithm [18]: consider the vertices in descending order according to the degree; we assign to a vertex the smallest available color not used by its neighbours, adding a fresh color if needed. In the DRD setting, deg is the maximal number of decision regions that any decision region can be overlapped with. In practice, greedy coloring needs much less colors than the upper bound. Thus DIRECT is potentially more efficient. In particular, when regions are disjoint, $\mathrm{deg} = 0$, and DIRECT reverts back to the EC^2 algorithm.

4 Experimental Results

We now consider three instances of the general non-myopic value of information problem. We compare DIRECT against several existing approaches. The first baseline is myopic optimization of the decision-theoretic value of information (VOI) [5]. At each step we greedily choose the test that maximizes the expected value given the current observations $\mathbf{x}_{\mathcal{A}}$, i.e., $t \in \arg\max_t \mathbb{E}_{x_t}\left[U(\mathbf{x}_{\mathcal{A}\cup\{x\}})\right]$. The second baseline is the recently proposed objective for addressing the DRD problem, HEC [3]. We also compare with algorithms designed for special cases of the DRD problem: GBS and EC^2. We compare with two versions of these algorithms: one with their original stopping criteria; and one with the stopping criteria of the DRD problem, which is referred to as GBS-DRD and EC^2-DRD.

Comparison-based preference learning. A comparison-based movie recommendation system [19] learns a user's movie preference (e.g., the favorable genre) by sequentially showing the user pairs of candidate movies, and letting her choose which one she prefers. We use the *MovieLens 100k* dataset [20], which consists a matrix of 1 to 5 ratings of 1682 movies from 943 users. For fair comparison with baselines, we adopt the same parameters as reported in [3]. That is, for each movie we extract a 10-d feature representation from the rating matrix through SVD. To generate decisions, we cluster movies using k-means, and assign each movie to the r closest cluster centers.



(a) *MovieLens* (b) *MovieLens* - time (c) *Robot* (d) *EMPCranes*

Figure 3: Experimental results

We demonstrate the performance of DIRECT on *MovieLens* in Figure 3a and 3b. We fix the number of clusters (i.e., decision regions) to 12, and vary r, the number of assigned regions for each hypothesis, from 1 to 6. Note that r controls the hyperedge cardinality in HEC, which crucially affects the computational complexity. As we can observe, while the *query complexity* (i.e., the number of queries needed till identifying the target region) of DIRECT is slightly higher than HEC (but universally lower than all other baselines), it is significantly faster to compute (for $r = 5$, HEC did not complete within a reasonable amount of time).

Active touch-based localization. Our second application is a robotic manipulation task of pushing a button, with uncertainty over the target's pose. We gather information with *guarded moves* [21], where the end effector moves along a path until contact is sensed. Those hypotheses which would not have produced contact at that location (e.g., they are far away) can be eliminated. Decisions correspond to putting the end effector at a particular location and moving forward. The coinciding decision region consists of all object poses where the button would successfully be pushed. Our goal is to concentrate all consistent hypotheses within a single decision region using the fewest tests.

We run DIRECT on both simulated data and a real robot platform. In the simulated experiments, we first sample an initial set of 20000 hypotheses, and then randomly generate decision regions, varying $|\mathcal{D}|$ while fixing $|\mathcal{T}| = 250$. Results are plotted in Figure 3c. Note that HEC cannot be computed in this experiment, as the overlap r becomes very large and HEC quickly becomes intractable. We see that DIRECT generally outperforms other baselines. Here, myopic VOI performs comparably – likely because the problem is solved within a short horizon.

Adaptive management for biodiversity conservation Our third application is a real-world value of information problem in natural resource management, where one needs to determine which management action should be undertaken for wild-life conservation. Specifically, the task is to preserve the *Eastern Migration Population of whooping cranes* (*EMP Cranes*). An expert panel came up with 8 hypotheses for possible causes of reproductive failure, along with 7 management strategies (as decisions). The decision-hypothesis utility matrix is specified in Table 5 of [4]. Tests aim to resolve specific sources of uncertainty. Our goal is to find the best conservation strategy using the minimal number of tests.

We assume that ε-optimal decisions are allowed for each hypothesis, where ε is the tolerance threshold. We further assume tests to be noisy, i.e., the test outcome of a particular hypothesis can be flipped. Maximally 1 flip is allowed for each outcome vector, which amounts to a total of 37 "noisy" hypotheses. When multiple hypotheses are consistent with a outcome vector, we assign the most probable one to that outcome. Results are plotted in Figure 3d. We see that HEC and DIRECT perform comparably well, while significantly outperforming myopic VOI and all other baselines.

5 Conclusion

We have proposed DIRECT, an efficient surrogate for the problem of nonmyopically optimizing value of information to achieve near-maximal utility. We prove that DIRECT is adaptive submodular, making it amenable for efficient greedy optimization. We demonstrated the efficiency and effectiveness of DIRECT extensively on three real-world applications, and showed that it compares favorably with existing approaches, while being significantly faster than competing methods. We believe that our results provide an important step towards solving challenging real-world information gathering problems.

References

[1] Igor Kononenko. Machine learning for medical diagnosis: history, state of the art and perspective. *Artificial Intelligence in Medicine*, 23:89–109, 2001.

[2] K. Chaloner and I. Verdinelli. Bayesian experimental design: A review. *Statistical Science*, 10(3):273–304, 1995.

[3] S. Javdani, Y. Chen, A. Karbasi, A. Krause, D. Bagnell, and S. Srinivasa. Near-optimal bayesian active learning for decision making. In *AISTATS*, 2014.

[4] M. C. Runge, S. J. Converse, and J. E. Lyons. Which uncertainty? using expert elicitation and expected value of information to design an adaptive program. *Biological Conservation*, 2011.

[5] R. A. Howard. Information value theory. In *IEEE Transactions on Systems Science and Cybernetics*, 1966.

[6] Andreas Krause and Carlos Guestrin. Optimal value of information in graphical models. *JAIR*, 35:557–591, 2009.

[7] D. Golovin, A. Krause, and D. Ray. Near-optimal bayesian active learning with noisy observations. *CoRR*, 2010.

[8] S. Dasgupta. Analysis of a greedy active learning strategy. In *NIPS*, 2004.

[9] M. Balcan, A. Beygelzimer, and J. Langford. Agnostic active learning. In *ICML*, 2006.

[10] Steve Hanneke. A bound on the label complexity of agnostic active learning. In *ICML*, 2007.

[11] B. Settles. *Active Learning*. Morgan & Claypool, 2012.

[12] V. T. Chakaravarthy, V. Pandit, S. Roy, P. Awasthi, and M. Mohania. Decision trees for entity identification: Approximation algorithms and hardness results. In *SIGMOD/PODS*, 2007.

[13] S. Rao Kosaraju, Teresa M. Przytycka, and Ryan S. Borgstrom. On an optimal split tree problem. In *WADS'99*, pages 157–168, 1999.

[14] D. Golovin, A. Krause, and D. Ray. Near-optimal bayesian active learning with noisy observations. In *Proc. Neural Information Processing Systems (NIPS)*, December 2010.

[15] D. Golovin and A. Krause. Adaptive submodularity: Theory and applications in active learning and stochastic optimization. *JAIR*, 2011.

[16] Andrew Guillory and Jeff Bilmes. Simultaneous learning and covering with adversarial noise. In *ICML*, pages 369–376, June 2011.

[17] A. Deshpande, L. Hellerstein, and D. Kletenik. Approximation algorithms for stochastic boolean function evaluation and stochastic submodular set cover. In *SODA*, 2014.

[18] D. J. Welsh and M. B. Powell. An upper bound for the chromatic number of a graph and its application to timetabling problems. *Computer Journal*, 1967.

[19] Amin Karbasi, Stratis Ioannidis, and Laurent Massoulie. Comparison-based learning with rank nets. In *ICML*, 2012.

[20] Jonathan L. Herlocker, Joseph A. Konstan, Al Borchers, and John Riedl. An algorithmic framework for performing collaborative filtering. In *SIGIR*, 1999.

[21] Peter M. Will and David D. Grossman. An experimental system for computer controlled mechanical assembly. *IEEE Trans. Computers*, 24(9):879–888, 1975.

[22] P.P. Wakker. *Prospect Theory: For Risk and Ambiguity*. Cambridge University Press, 2010.

[23] Amos Tversky and Daniel Kahneman. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4), 1992.

[24] William F. Sharpe. Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *The Journal of Finance*, 1964.

[25] Debajyoti Ray, Daniel Golovin, Andreas Krause, and Colin Camerer. Bayesian rapid optimal adaptive design (broad): Method and application distinguishing models of risky choice. *Tech. Report*, 2012.

A Table of Notations Defined in the Main Paper

We summarize the notations used in the main paper in Table 1.

Table 1: A reference table of notations used in the main paper

n	total number of tests
N	total number of hypotheses in the DRD problem
m	total number of decision regions
r	number of EC^2 instances needed after applying graph coloring
k	bounded number of label flips in the restricted noise model
$\mathcal{T}$	set of all available tests
$\mathcal{A}$	subset of tests
t	test
$\mathcal{X}$	domain of test observations
X_t	(observable) random variable associated with a test t
x_t	observed value of a test t
$\mathbf{x}_{\mathcal{A}}$	vector of observations of tests in $\mathcal{A}$
$\mathbf{x}_{\mathcal{T}}$	vector of observations of all tests
$\mathbb{P}[\mathbf{x}_{\mathcal{T}}]$	probability of a specific realization
p_{min}	the minimum prior probability of any set of observations
$\mathcal{Y}$	domain of the hidden states
Y	random variable associated with a hidden state
y	value of the hidden state Y
$\mathcal{D}$	set of decisions that can be made
d	decision
$\mathcal{R}_d$	the decision region indexed by d
$u(d, y)$	utility function quantifying the benefit of making a decision $d \in \mathcal{D}$ for any $y \in \mathcal{Y}$
$U(d \mid \mathbf{x}_{\mathcal{A}})$	the expected value of a decision d after observing $\mathbf{x}_{\mathcal{A}}$
$\text{VoI}(\mathbf{x}_{\mathcal{A}})$	the value of a specific set of observations $\mathbf{x}_{\mathcal{A}}$
$R(d \mid \mathbf{x}_{\mathcal{A}})$	the regret of a decision d given observations $\mathbf{x}_{\mathcal{A}}$
ε	tolerance, maximal regret allowed
π	policy, i.e., a partial mapping from observation vectors to tests
$\mathcal{S}(\pi, \mathbf{x}_{\mathcal{T}})$	the set of observations obtained by playing policy policy π, under realization $x_{\mathcal{T}}$
$c(t)$	the cost of performing a test $t \in \mathcal{T}$
$c(\mathbf{x}_{\mathcal{A}})$	the cost of performing a sequence of tests $\mathbf{x}_{\mathcal{A}} \in \mathcal{X}^{\mathcal{A}}$
$\text{cost}(\pi)$	the expected cost of a policy π
$\mathcal{H}$	set of hypotheses in the DRD problem
h	hypothesis
H	a random variable distributed over $\mathcal{H}$.
$f_t(h)$	the realization of test t under hypothesis h
$G = (V, E)$	EC^2 graph
$w(h, h')$	weight of edge $(h, h') \in E$ in the EC^2 graph G
f_{EC}	the EC^2 objective function
$\Delta_{f_{EC}}$	the expected marginal benefit in f_{EC} by adding test t to $\mathbf{x}_{\mathcal{A}}$
f_{DRD}	the DiRECt objective function
$\Delta_{f_{DRD}}(t \mid \mathbf{x}_{\mathcal{A}})$	the expected marginal benefit in f_{DRD} by adding test t to $\mathbf{x}_{\mathcal{A}}$
$\mathcal{G} = \{\mathcal{D}, \mathcal{E}\}$	(undirected) graph over $\mathcal{D}$; edges are drawen between overlapped decision regions
deg	the maximal degree of $\mathcal{G}$
Θ	latent (nuisance) variable that models the noise of a hidden state
θ	value of latent variable Θ
$\delta(h, \hat{h})$	the total number of label (i.e, test outcome) flips from h to $\hat{h}$

B Efficient Computation of DIRECT

Essentially, DIRECT is built upon a collection of EC^2 objectives. In this section, we show that EC^2 (and thus DIRECT) can be computed in *linear* time in the number of hypotheses.

Recall the definition of edge weight in EC^2: $w(\{h, h'\}) = \mathbb{P}[h] \cdot \mathbb{P}[h']$. Let $\mathbb{P}[\mathcal{R}_i]$ be the total prior probability mass of all hypotheses h in $\mathcal{R}_i$. Then the weight of edges between distinct (disjoint) decision regions $\mathcal{R}_i, \mathcal{R}_j$ is

$$w(\mathcal{R}_i \times \mathcal{R}_j) = \sum_{h \in \mathcal{R}_i, h' \in \mathcal{R}_j} \mathbb{P}[h]\,\mathbb{P}[h'] = \Big(\sum_{h \in \mathcal{R}_i} \mathbb{P}[h]\Big)\Big(\sum_{h' \in \mathcal{R}_j} \mathbb{P}[h']\Big) = \mathbb{P}[\mathcal{R}_i]\,\mathbb{P}[\mathcal{R}_j],$$

and the total weight is $\frac{1}{2}\sum_{i \neq j} w(\mathcal{R}_i \times \mathcal{R}_j) = \frac{1}{2}\Big(\sum_i \mathbb{P}[\mathcal{R}_i]\Big)^2 - \sum_i \mathbb{P}[\mathcal{R}_i]^2 = \frac{1}{2}\big(1 - \sum_i \mathbb{P}[\mathcal{R}_i]^2\big)$.

Since the EC^2 objective (i.e., the weight reduction) could be computed as the total weight of edges subtracting the remaining weight[3], we can thus compute it efficiently through the following equation [7]:

$$f_{EC}(\mathbf{x}_\mathcal{A}) = \frac{1}{2}\Big[1 - \sum_i \mathbb{P}[\mathcal{R}_i]^2 + \sum_i \mathbb{P}[\mathcal{R}_i \cap \mathcal{H}(\mathbf{x}_\mathcal{A})]^2 - \Big(\sum_i \mathbb{P}[\mathcal{R}_i \cap \mathcal{H}(\mathbf{x}_\mathcal{A})]\Big)^2\Big],$$

where $\mathbb{P}[\mathcal{R}_i \cap \mathcal{H}(\mathbf{x}_\mathcal{A})]$ is the mass of all hypotheses h in $\mathcal{R}_i$ consistent with observations $\mathbf{x}_\mathcal{A}$.

C Proofs

C.1 Proof of Lemma 1

In the following, we show that the function defined in the form of Eq. 3 is strongly adaptive monotone and adaptive submodular.

Proof of Lemma 1. We first show f_{DRD} is strongly adaptively monotone: We know that each individual f^i_{EC} is strongly adaptively monotone. Moreover, the partial derivative of f_{DRD} w.r.t. each f^i_{EC} is non-negative. Applying the chain rule of derivatives, we know that f_{DRD} is strongly adaptively monotone.

To proof adaptive submodularity, we need to prove that for all $\mathbf{x}_\mathcal{A} \preceq \mathbf{x}_\mathcal{B}$ and $t \in \mathcal{T}$, it holds that $\Delta_{f_{DRD}}(t \mid \mathbf{x}_\mathcal{A}) \geq \Delta_{f_{DRD}}(t \mid \mathbf{x}_\mathcal{B})$. First we introduce several auxiliary notations, as shown in Table 2. Let $n_a(\mathbf{x}_\mathcal{A}) = \sum_i n_{i,a}(\mathbf{x}_\mathcal{A})$ be the number of hypotheses in the current hypotheses space given $\mathbf{x}_\mathcal{A}$ and $X_t = a$, and $n_\mathcal{T}(\mathbf{x}_\mathcal{A}) = |\mathcal{H}(\mathbf{x}_\mathcal{A})|$ be the number of hypotheses that are consistent with the observation $\mathbf{x}_\mathcal{A}$ (See Table 2 for a list of notations used in this proof).

Table 2: A reference table of auxiliary notations

$n_\mathcal{T}(\mathbf{x}_\mathcal{A})$	$\lvert\mathcal{H}(\mathbf{x}_\mathcal{A})\rvert$, the number of hypotheses that are consistent with the observation $\mathbf{x}_\mathcal{A}$.
$n_i(\mathbf{x}_\mathcal{A})$	$\lvert\mathcal{H}(\mathbf{x}_\mathcal{A}) \cap \mathcal{R}_i\rvert$, the number of hypotheses in $\mathcal{R}_i$ that are consistent with $\mathbf{x}_\mathcal{A}$.
$n_a(\mathbf{x}_\mathcal{A})$	$\sum_i n_{i,a}(\mathbf{x}_\mathcal{A})$, the number of hypotheses in the current version space given $\mathbf{x}_\mathcal{A}$ and $X_t = a$.
$n_{i,a}(\mathbf{x}_\mathcal{A})$	$\lvert\{h : h \in \mathcal{H}(\mathbf{x}_\mathcal{A}, X_t = a) \cap \mathcal{R}_i\}\rvert$, the number of hypotheses in $\mathcal{R}_i$ that are consistent with the observation $\mathbf{x}_\mathcal{A}$ and $X_t = a$.
$\mathbf{n}(\mathbf{x}_\mathcal{A})$	the vector consisting of $n_{i,a}(\mathbf{x}_\mathcal{A})$ for all i and a.
ϕ	the expected marginal benefit of a test given some observations.

[3]Essentially, EC^2 is efficiently computed as elementary symmetric polynomials. The general strategy is to compute the sum of all edge weights between hypotheses, and then subtract those that share a region. The same technique is also used in [3] for efficient implementation of HEC. We refer interested reader to [3] for more details.

As of [7], we can represent the marginal gain of f_{EC} on each graph as a function $\phi(\cdot)$ only depending on $\mathbf{n}(\mathbf{x}_{\mathcal{A}})$:

$$\Delta_{f_{EC}}(t \mid \mathbf{x}_{\mathcal{A}}) = \phi(\mathbf{n}(\mathbf{x}_{\mathcal{A}})) = \frac{1}{2}\sum_{i\neq j}\sum_{a\neq b} n_{i,a}(\mathbf{x}_{\mathcal{A}}) \cdot n_{j,b}(\mathbf{x}_{\mathcal{A}}) + \sum_{a}\frac{n_a}{n_{\mathcal{T}}} \cdot \frac{1}{2}\sum_{i\neq j}\sum_{b\neq a} n_{i,b} \cdot n_{j,b} \tag{5}$$

Now let $n_{k,c}$ be the number of hypotheses in *auxiliary* equivalence class k, which are consistent with the observation $X_t = c$. From [7], we get $\partial\phi/\partial n_{k,c} \geq 0$ for any choice of k and c.

To show that $\Delta_{f_{DRD}}(t \mid \mathbf{x}_{\mathcal{A}}) = \phi_{f_{DRD}}(\mathbf{n}(\mathbf{x}_{\mathcal{A}}))$ is monotone decreasing with more observations, we need to show that for any k and c, it holds that $\partial\phi_{f_{DRD}}(\mathbf{n}(\mathbf{x}_{\mathcal{A}}))/\partial n_{k,c} \geq 0$. Denote the set $\mathcal{A} \cup \{t\}$ as $\mathcal{A} + t$. By the definition of $\Delta(t \mid \mathbf{x}_{\mathcal{A}})$, we know

$$\begin{aligned}
&\Delta_{f_{DRD}}(t \mid \mathbf{x}_{\mathcal{A}}) \\
=&\mathbb{E}\left[\left(1 - \prod_{i}^{m}\left(1 - f_{EC}^{i}(\mathbf{x}_{\mathcal{A}+t})\right)\right) - \left(1 - \prod_{i}^{m}\left(1 - f_{EC}^{i}(\mathbf{x}_{\mathcal{A}})\right)\right)\right] \\
=&\mathbb{E}\left[(1 - f_{EC}^{1}(\mathbf{x}_{\mathcal{A}})) \cdot \prod_{i\neq 1}^{m}\left(1 - f_{EC}^{i}(\mathbf{x}_{\mathcal{A}})\right) - (1 - f_{EC}^{1}(\mathbf{x}_{\mathcal{A}+t})) \cdot \prod_{i\neq 1}^{m}\left(1 - f_{EC}^{i}(\mathbf{x}_{\mathcal{A}+t})\right)\right]
\end{aligned} \tag{6}$$

We first show for the simple case, where there are only two regions, the objective $f_{EC}^{(2)}$ is adaptive submodular w.r.t. uniform priors. For discussion simplicity we drop the normalization constants Q_i from the analysis.

Define $\delta_i(x_t \mid \mathbf{x}_{\mathcal{A}}) = f_{EC}^{i}(\mathbf{x}_{\mathcal{A}+t}) - f_{EC}^{i}(\mathbf{x}_{\mathcal{A}})$. If there are two regions, i.e., $m = 2$, Eq 6 becomes

$$\begin{aligned}
&\Delta_{EC}(t \mid \mathbf{x}_{\mathcal{A}}) \\
=&\mathbb{E}\left[(1 - f_{EC}^{1}(\mathbf{x}_{\mathcal{A}})) \cdot (1 - f_{EC}^{2}(\mathbf{x}_{\mathcal{A}})) - (1 - f_{EC}^{1}(\mathbf{x}_{\mathcal{A}+t})) \cdot (1 - f_{EC}^{2}(\mathbf{x}_{\mathcal{A}+t}))\right] \\
=&\mathbb{E}\left[f_{EC}^{1}(\mathbf{x}_{\mathcal{A}+t}) - f_{EC}^{1}(\mathbf{x}_{\mathcal{A}}) + f_{EC}^{2}(\mathbf{x}_{\mathcal{A}+t}) - f_{EC}^{2}(\mathbf{x}_{\mathcal{A}}) - \left(f_{EC}^{1}(\mathbf{x}_{\mathcal{A}+t})f_{EC}^{2}(\mathbf{x}_{\mathcal{A}+t}) - f_{EC}^{1}(\mathbf{x}_{\mathcal{A}})f_{EC}^{2}(\mathbf{x}_{\mathcal{A}})\right]\right. \\
=&\mathbb{E}\left[\delta_1(x_t \mid \mathbf{x}_{\mathcal{A}}) + \delta_2(x_t \mid \mathbf{x}_{\mathcal{A}}) - (\delta_1(x_t \mid \mathbf{x}_{\mathcal{A}})f_{EC}^{2}(\mathbf{x}_{\mathcal{A}+t}) + \delta_2(x_t \mid \mathbf{x}_{\mathcal{A}})f_{EC}^{1}(\mathbf{x}_{\mathcal{A}})) \mid \mathbf{x}_{\mathcal{A}}\right] \\
=&\mathbb{E}\left[(1 - f_{EC}^{1}(\mathbf{x}_{\mathcal{A}}))\delta_2(x_t \mid \mathbf{x}_{\mathcal{A}}) \mid \mathbf{x}_{\mathcal{A}}\right] + \mathbb{E}\left[(1 - f_{EC}^{2}(\mathbf{x}_{\mathcal{A}+t}))\delta_1(x_t \mid \mathbf{x}_{\mathcal{A}}) \mid \mathbf{x}_{\mathcal{A}}\right] \\
=&\left(1 - f_{EC}^{1}(\mathbf{x}_{\mathcal{A}})\right)\mathbb{E}\left[\delta_2(x_t \mid \mathbf{x}_{\mathcal{A}}) \mid \mathbf{x}_{\mathcal{A}}\right] + \mathbb{E}\left[\left(1 - f_{EC}^{2}(\mathbf{x}_{\mathcal{A}+t})\right)\delta_1(x_t \mid \mathbf{x}_{\mathcal{A}}) \mid \mathbf{x}_{\mathcal{A}}\right]
\end{aligned} \tag{7}$$

For the first term on the R.H.S. of Eq. 7, we have

$$\left(1 - f_{EC}^{1}(\mathbf{x}_{\mathcal{A}})\right)\mathbb{E}\left[\delta_2(x_t \mid \mathbf{x}_{\mathcal{A}}) \mid \mathbf{x}_{\mathcal{A}}\right] \geq \left(1 - f_{EC}^{1}(\mathbf{x}_{\mathcal{B}})\right)\mathbb{E}\left[\delta_2(x_t \mid \mathbf{x}_{\mathcal{B}}) \mid \mathbf{x}_{\mathcal{B}}\right] \tag{8}$$

Let the second term be $\theta(\mathbf{n})$, and denote $h(\mathbf{n}) = 1 - f_{EC}^{2}(\mathbf{x}_{\mathcal{A}+t})$. In the following, we will show that $\partial\theta(\mathbf{n})/\partial n_{k,c} \geq 0$ for all $n_{k,c}$.

$$\begin{aligned}
\theta(\mathbf{n}) &= \mathbb{E}\left[h(\mathbf{n})\delta_1(x_t \mid \mathbf{x}_{\mathcal{A}}) \mid \mathbf{x}_{\mathcal{A}}\right] \\
&= \sum_{a} h(\mathbf{n})\frac{n_a}{n_{\mathcal{T}}} \cdot \frac{1}{2}\left\{\sum_{i\neq j}\sum_{b\neq d} n_{i,b}(\mathbf{x}_{\mathcal{A}}) \cdot n_{j,d}(\mathbf{x}_{\mathcal{A}}) + \sum_{i\neq j}\sum_{b\neq a} n_{i,b}(\mathbf{x}_{\mathcal{A}}) \cdot n_{j,b}(\mathbf{x}_{\mathcal{A}})\right\}
\end{aligned}$$

Taking the partial derivative of $\theta(\mathbf{n})$ w.r.t. $n_{k,c}$, we have

$$\begin{aligned}
\frac{\partial\theta(\mathbf{n})}{\partial n_{k,c}} =& \sum_{a}\frac{\partial h(\mathbf{n})}{\partial n_{k,c}} \cdot \frac{n_a}{n_{\mathcal{T}}} \cdot \frac{1}{2}\left\{\sum_{i\neq j}\sum_{b\neq d} n_{i,b}(\mathbf{x}_{\mathcal{A}}) \cdot n_{j,d}(\mathbf{x}_{\mathcal{A}}) + \sum_{i\neq j}\sum_{b\neq a} n_{i,b}(\mathbf{x}_{\mathcal{A}}) \cdot n_{j,b}(\mathbf{x}_{\mathcal{A}})\right\} \\
&+ \sum_{a} h(\mathbf{n}) \cdot \frac{\partial}{\partial n_{k,c}}\left\{\frac{n_a}{2n_{\mathcal{T}}} \cdot \sum_{i\neq j}\sum_{b\neq d} n_{i,b}(\mathbf{x}_{\mathcal{A}}) \cdot n_{j,d}(\mathbf{x}_{\mathcal{A}}) + \frac{n_a}{2n_{\mathcal{T}}} \cdot \sum_{i\neq j}\sum_{b\neq a} n_{i,b}(\mathbf{x}_{\mathcal{A}}) \cdot n_{j,b}(\mathbf{x}_{\mathcal{A}})\right\}
\end{aligned} \tag{9}$$

Since $f_{EC}^{2}(\mathbf{x}_{\mathcal{A}+t})$ is monotone decreasing w.r.t. $n_{k,c}$, $h(\mathbf{n})$ is monotone increasing, and thus $\partial h/\partial n_{k,c} \geq 0$. Therefore, the first term on the R.H.S. of Eq. 9 is nonnegative.

Let $p = \frac{1}{2}\sum_{i \neq j, b \neq d} n_{i,b} n_{j,d}$, and $q_a = \frac{1}{2}\sum_{i \neq j, b \neq a} n_{i,b} n_{j,b}$. For simplicity we drop the dependency of variables on $\mathbf{x}_\mathcal{A}$. Then the second term on the R.H.S. of Eq. 9 is

$$
\begin{aligned}
&\sum_a h(\mathbf{n}) \cdot \frac{\partial}{\partial n_{k,c}} \left\{ n_a \cdot \frac{1}{n_\mathcal{T}} \cdot p + n_a \cdot \frac{1}{n_\mathcal{T}} \cdot q_a \right\} \\
=& h(\mathbf{n}) \cdot \underbrace{\frac{\partial}{\partial n_{k,c}} \left\{ n_c \cdot \frac{1}{n_\mathcal{T}} \cdot p + n_c \cdot \frac{1}{n_\mathcal{T}} \cdot q_c \right\}}_{\textcircled{1}} + \sum_{a \neq c} h(\mathbf{n}) \cdot \underbrace{\frac{\partial}{\partial n_{k,c}} \left\{ n_a \cdot \frac{1}{n_\mathcal{T}} \cdot p + n_a \cdot \frac{1}{n_\mathcal{T}} \cdot q_a \right\}}_{\textcircled{2}}
\end{aligned}
\tag{10}
$$

Expand term $\textcircled{1}$ to get

$$
\begin{aligned}
\textcircled{1} &= \frac{n_c}{n_\mathcal{T}} \cdot \frac{\partial p}{\partial n_{k,c}} + \frac{p}{n_\mathcal{T}} \cdot \cancelto{1}{\frac{\partial n_c}{\partial n_{k,c}}} + p n_c \cdot \frac{\partial (1/n_\mathcal{T})}{\partial n_{k,c}} + \frac{n_c}{n_\mathcal{T}} \cdot \cancelto{0}{\frac{\partial q_c}{\partial n_{k,c}}} + \frac{q_c}{n_\mathcal{T}} \cdot \cancelto{1}{\frac{\partial n_c}{\partial n_{k,c}}} + q_c n_c \cdot \frac{\partial (1/n_\mathcal{T})}{\partial n_{k,c}} \\
&= \frac{n_c}{n_\mathcal{T}} \cdot \sum_{j \neq k, b \neq c} n_{j,b} + \frac{p}{n_\mathcal{T}} - \frac{p n_c}{n_\mathcal{T}^2} + \frac{q_c}{n_\mathcal{T}} - \frac{q_c n_c}{n_\mathcal{T}^2} \\
&= \frac{n_c}{n_\mathcal{T}} \cdot \sum_{j \neq k, b \neq c} n_{j,b} + p \cdot \left(\frac{1}{n_\mathcal{T}} - \frac{n_c}{n_\mathcal{T}^2} \right) + q_c \cdot \left(\frac{1}{n_\mathcal{T}} - \frac{n_c}{n_\mathcal{T}^2} \right) \geq 0
\end{aligned}
\tag{11}
$$

Similarly, for term $\textcircled{2}$,

$$
\begin{aligned}
\textcircled{2} &= \frac{n_a}{n_\mathcal{T}} \cdot \underbrace{\frac{\partial p}{\partial n_{k,c}}}_{\sum_{j \neq k, b \neq c} n_{j,b}} + \frac{p}{n_\mathcal{T}} \cdot \cancelto{0}{\frac{\partial n_a}{\partial n_{k,c}}} + p n_a \cdot \frac{\partial (1/n_\mathcal{T})}{\partial n_{k,c}} + \frac{n_a}{n_\mathcal{T}} \cdot \underbrace{\frac{\partial q_a}{\partial n_{k,c}}}_{\sum_{j \neq k} n_{j,c}} + \frac{q_a}{n_\mathcal{T}} \cdot \cancelto{0}{\frac{\partial n_a}{\partial n_{k,c}}} + q_a n_a \cdot \frac{\partial (1/n_\mathcal{T})}{\partial n_{k,c}} \\
&= \frac{n_a}{n_\mathcal{T}} \cdot \sum_{j \neq k, b \neq c} n_{j,b} - \frac{p n_a}{n_\mathcal{T}^2} + \frac{n_a}{n_\mathcal{T}} \cdot \sum_{j \neq k} n_{j,c} - \frac{q_a n_a}{n_\mathcal{T}^2} \\
&= \frac{n_a}{n_\mathcal{T}} \cdot \left\{ \sum_{j \neq k} \sum_b n_{j,b} - \underbrace{\left(\frac{p}{n_\mathcal{T}} + \frac{q_a}{n_\mathcal{T}} \right)}_{\textcircled{3}} \right\}
\end{aligned}
\tag{12}
$$

Substitute $p = \frac{1}{2}\sum_{i \neq j, b \neq d} n_{i,b} n_{j,d}$, and $q_a = \frac{1}{2}\sum_{i \neq j, b \neq a} n_{i,b} n_{j,b}$ in term $\textcircled{3}$ to get:

$$
\begin{aligned}
\frac{p}{n_\mathcal{T}} + \frac{q_a}{n_\mathcal{T}} &= \frac{1}{2} \sum_{i \neq j, b \neq d} n_{i,b} \frac{n_{j,d}}{n_\mathcal{T}} + \frac{1}{2} \sum_{i \neq j, b \neq a} n_{i,b} \frac{n_{j,b}}{n_\mathcal{T}} \\
&\leq \frac{1}{n_\mathcal{T}} \cdot \frac{1}{2} \sum_{i \neq j, b \neq d} \left(n_{i,b} n_{j,d} + n_{i,b} n_{j,b} \right) \\
&\leq \frac{1}{n_\mathcal{T}} \left(\sum_{i,d} n_{i,d} \right) \cdot \left(\sum_{j \neq k} \sum_b n_{j,b} \right) \\
&= \sum_{j \neq k} \sum_b n_{j,b}
\end{aligned}
\tag{13}
$$

Hence term $\textcircled{2}$ is nonnegative. Combining Eq. 10 to 13 with Eq. 9, we get $\partial \theta(\mathbf{n}) / \partial n_{k,c} \geq 0$. Therefore, fix $\mathbf{x}_\mathcal{A} \preceq \mathbf{x}_\mathcal{B}$ and $t \in \mathcal{T}$, it holds that $\Delta_{ER}(t \mid \mathbf{x}_\mathcal{A}) \geq \Delta_{ER}(t \mid \mathbf{x}_\mathcal{B})$ for the case where there are two regions, and thus f_{EC} is adaptive submodular for $m = 2$ w.r.t. a uniform prior (note that we can adapt the proof technique from [7] to prove A.S. for arbitrary prior).

Now assume that $f_{DRD}^{(m)}$ is adaptive submodular for $m = k$ and $k > 2$, and we want to prove when $m = k + 1$, $f_{DRD}^{(k+1)}$ is also adaptive submodular. By definition, we have

$$
\begin{aligned}
f_{DRD}^{(k+1)} &= 1 - \prod_{i=1}^{k+1} \left(1 - f_{EC}^{i}(\mathcal{S}(\pi, \mathbf{x}_{\mathcal{T}}))\right) \\
&= 1 - (1 - f_{EC}^{k+1}(\mathcal{S}(\pi, \mathbf{x}_{\mathcal{T}})) \cdot \prod_{i=1}^{k} \left(1 - f_{EC}^{i}(\mathcal{S}(\pi, \mathbf{x}_{\mathcal{T}}))\right) \\
&= 1 - (1 - f_{EC}^{k+1}(\mathcal{S}(\pi, \mathbf{x}_{\mathcal{T}})) \cdot (1 - f_{DRD}^{(k)})
\end{aligned}
$$

Since $f_{DRD}^{(k)}$ is adaptive submodular and strongly adaptive monotone, we can apply the same analysis for the two region case, to the above problem. Therefore, $f_{DRD}^{(k+1)}$ is adaptive submodular, and thus $f_{DRD}^{(m)}$ is adaptive submodular for any $m \geq 1$. □

Remarks ("intuitive explanation" of the proof of Lemma 1). In fact, one can find concrete examples where Noisy-OR does not preserve adaptive submodularity. Fortunately, for EC^2-like objectives, we have proved that it does preserve adaptive submodularity. The intuition lies in that the EC^2 objective characterizes a class of adaptive submodular functions with certain structures, which offers enough slack for our proof to go through.

C.2 Proof of Theorem 2

Proof. Let Q be the quota to be achieved, and η be any value such that $f_{DRD}(\mathcal{S}(\pi, \mathbf{x}_{\mathcal{T}})) > Q - \eta$ implies $f_{DRD}(\mathcal{S}(\pi, \mathbf{x}_{\mathcal{T}})) = Q$, then by Theorem 10 of [15], the cost of π_{DRD} satisfies

$$c(\pi_{DRD}) \leq c(\pi^*)(\ln(Q/\eta) + 1).$$

In our case, apply $Q = 1$ and $\eta \geq \left(\frac{1}{p_{min}^2}\right)^m$ to get $c(\pi_{DRD}) \leq \text{cost}(\pi^*)(2m \ln(1/p_{min}) + 1)$. □

D Supplemental Experiments

In this section, we provide necessary implementation details that could be used to regenerate the results for the applications discussed in Section 4. Moreover, we test DiRECt on a new application: (1) preference elicitation in behavioral economics, where we want to adaptively decide which theory best explains observed risky choices. Our experimental results demonstrate consistent results that DiRECt is effective in various decision making tasks.

D.1 Implementation Details and Supplemental Results

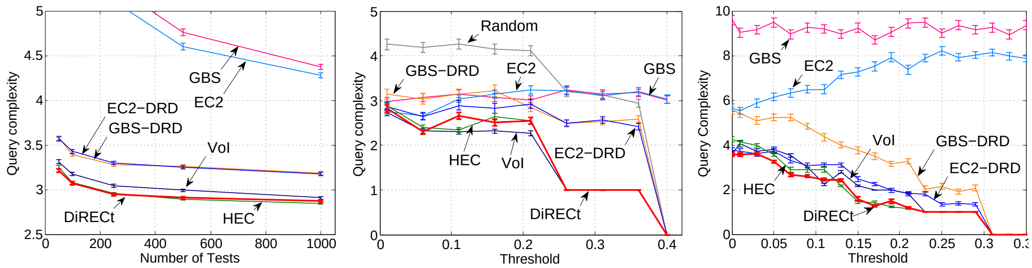
Table 3 summarizes how the applications discussed in this paper fit into our Decision Region Determination framework. The fourth application (risky choice selection) is introduced in Section D.2.

Application	**Movie recomm..**	**Bio. conservation**	**Touch-based Localization**	**Risky choice selection**
Test	pair of movies	monitoring / probing	guarded move	pair of lottery choices
Hidden Var	target movie	cause for nest failure	target location	parametrized theory
Decision	recommendation	conservation action	manipulation action	theory adoption

Table 3: Overview of how different applications discussed in this paper fit into our framework

D.1.1 Comparison-based Preference Learning

This section explains the detailed experimental setup for the preference learning application. Imagine that we want to learn a user's preferred category of movies. The user will be happy as long as any movie in this genre is recommended to her. After showing the user a sequence of candidate movie



(a) *Robot* - varying $|\mathcal{T}|$ (b) *EMPCrances* - noise-free (c) *Risky Choice Theory*

Figure 4: Supplemental experimental results.

pairs (i.e., *tests*), we can get a set of feedbacks (i.e., a vector outcomes that align with the user's movie preference – the (ground truth) *hypothesis*). After receiving feedback from each test, we remove the movies that we believe do not reflect user's interest (e.g., movies that are more similar with the one that the user chooses to dislike). Further suppose that we have a pool of candidate movies to recommend. Once all the remaining movies in our pool are in the same category (i.e., *decision region*), we can recommend any of the movies to the user. Our goal is identify such a category by asking as few pair-wise comparison questions as possible.

Constructing decision regions To measure the similarity of candidate movies, we extract movie features by computing a low-rank approximation of the user/rating matrix of the *MoiveLens 100k* dataset through singular value decomposition (SVD). Specifically, we extract a 10-dimensional feature vector for each movie. We then use k-means to partition the set of movies into r (non-overlapping) clusters in the Euclidean space, corresponding to decision regions. We choose the k-means cluster centers as the centroids representing the "categories" that a user may be interested in. Since one movie can usually belong to several categories, we assign each movie to the category that is represented by the closest centroids, giving us overlapped decision regions.

Generating tests As mentioned earlier, tests are pairs of movies. We generate a set of ≈1.4 million tests from the 1682 movies in the *MoiveLens 100k* dataset. Usually, to distinguish a movie from the rest of the pool, we don't need to perform all the tests. Rather, we want to extract a subset of tests from all the available tests, such that by performing this subset of tests, one can uniquely distinguish all the movies in our pool. To select this subset, we first build a binary matrix $A = \{a_{i,j}\}_{1682\times1682}$ of size 1682×1682, representing all pairs of movies to be distinguished. If performing a test $t = (m_1, m_2)$ can distinguish a pair of target movies indexed by(i, j) (meaning that by performing the test, one can tell which one of (i, j) is more favorable), then we fill the entry $a_{i,j} = 1$, indicating that we can distinguish i from j by performing this test. Therefore, we start from an empty set T of tests, and keep adding tests (following some random order) into the T till the matrix A is filled up. This amounts to a total number ≈ 100 tests, with which we can uniquely identify any of the 1682 movies.

D.1.2 Active Touch-based Localization

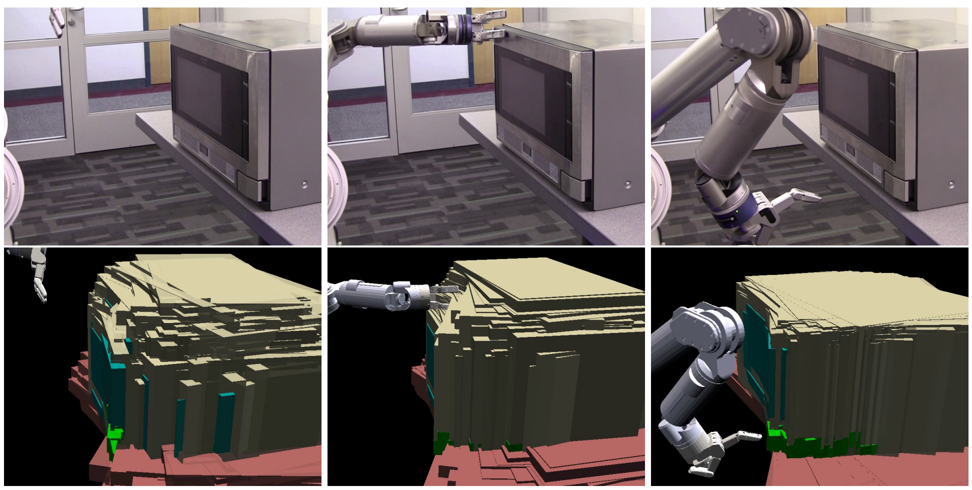
In this section, we provide implementation details and further results on the active touch-based localization application.

Generating hypotheses To model the pose uncertainty of the target, we use 4 parameters: (x, y, z) for positional uncertainty, and θ for rotation about the z axis. An initial set of 20000 hypotheses are sampled from a normal distribution $N(\mu, \Sigma)$, where μ is some initial location (e.g., from a camera), and Σ is diagonal with $\sigma_x = \sigma_y = \sigma_z = 2.5$cm, and $\sigma_\theta = 7.5°$.

Comparing DiRECt *with* HEC In results included in the main paper, it's prohibitive to run HEC, because the overlap between regions are large. As supplemental results, we also want to compare DiRECt with HEC, on problem instances where HEC can practically run. To ensure that, we preselect a grid of 25 button pushing actions $\mathcal{D}$ while ensuring the overlap r is minimal, so that the HEC objective can be computed in reasonable time. Note that to run DIRECT, we don't need to enforce such strict constraints. We randomly generate guarded moves $\mathcal{T}$ to select from. To compute the *myopic value of information* (VOI) [5], we define a utility function $u(h, \mathcal{R})$ which is 1 if $h \in \mathcal{R}$ and 0 otherwise. In Figure 4a we show the number of test guarded moves needed for different algorithms, when varying $|\mathcal{T}|$. As we can observe from the results, DIRECT performs essentially the same as HEC on this problem instance, while slightly outperforms VOI. We would expect that for longer horizons, myopic VOI would not perform as well.

Demonstrating DIRECT *on a real robot platform* In the supplementary video, we demonstrate DIRECT on a real robot manipulation task, where the goal is to push a button with the finger of a robotic end effector. We can see that the microwave button is successfully localized on the fourth touch attempt – when all consistent hypotheses are encapsulated by one single decision region (also see Figure 5 for more details).

We also demonstrate DIRECT on a real robot platform as illustrated in Figure 5. See supplemental material for more results and a video demonstration.



(a) Hypotheses (b) Tests (c) Decision Regions

Figure 5: Experimental setup for touch-based localization. (a) Uncertainty is represented by hypotheses over object pose. (b) Tests are guarded moves, where the end effector moves along a path until contact is sensed. Hypotheses which could not have produced contact at that location (e.g. they are too far or too close) are removed. (c) Decisions are button-push attempts: trajectories starting at a particular location, and moving forward. The corresponding region consists of all poses for which that button push would succeed.

D.1.3 Adaptive Management for Biodiversity Conservation

In the main paper, we have shown how DIRECT compares with baselines when the test outcomes are noisy. Figure 4b shows the results on clean data where no flip of test outcome is allowed (i.e., 7 hypotheses, 7 tests, 8 decision regions). We can see that DIRECT performs comparably with the baselines. When tolerance parameter is large enough, i.e., $\varepsilon \geq 0.4$, all initial decisions are "near"-optimal, in which case we don't need to perform any tests, and therefore query complexity becomes 0.

D.2 Supplemental Experiment: Preference Elicitation in Behavioral Economics.

We further conduct experiments in an experimental design task. Several theories have been proposed in behavioral economics to explain how people make decisions under risk and uncertainty. We test DIRECT on six theories of subjective valuation of risky choices [22, 23, 24], namely the (1) *expected utility with constant relative risk aversion*, (2) *expected value*, (3) *prospect theory*, (4) *cumulative prospect theory*, (5) *weighted moments*, and (6) *weighted standardized moments*. Choices are between risky lotteries, i.e., known distribution over payoffs (e.g., the monetary value gained or lost). Tests are pairs of lotteries, and hypotheses correspond to parametrized theories that predict, for a given test, which lottery is preferable. The goal, is to adaptively select a sequence of tests to present to a human subject in order to distinguish which of the six theories best explains the subject's responses.

We employ the same set of parameters used in [25] to generate tests and hypotheses. The original setup in [25] was designed for testing EC^2, and therefore test realizations of different theories cannot collide. In our experiments, we allow a tolerance ϵ - that is, if one hypothesis differs from another by at most ε, they are considered to be similar, and thus have the same set of optimal decisions. Results for simulated test outcomes with varying ε are shown in Figure 4c. We see that DIRECT performs best in this setting.