Combinatorial Selection with Costly Information

Shuchi Chawla* Dimitris Christou† Amit Harlev† Ziv Scully‡

Abstract

We consider a class of optimization problems over stochastic variables where the algorithm can learn information about the value of any variable through a series of costly steps; we model this information acquisition process as a Markov Decision Process (MDP). The algorithm's goal is to minimize the cost of its solution plus the cost of information acquisition, or alternately, maximize the value of its solution minus the cost of information acquisition. Such *bandit superprocesses* have been studied previously but solutions are known only for fairly restrictive special cases.

We develop a framework for approximate optimization of bandit superprocesses that applies to arbitrary acyclic MDPs with a matroid feasibility constraint. Our framework establishes a bound on the optimal cost through a novel cost amortization; it then couples this bound with a notion of local approximation that allows approximate solutions for each component MDP in the superprocess to be composed without loss into a global approximation. We use this framework to obtain approximately optimal solutions for several variants of bandit superprocesses for both maximization and minimization. We obtain new approximations for combinatorial versions of the previously studied Pandora's Box with Optional Inspection and Pandora's Box with Partial Inspection; the less-studied Additive Pandora's Box problem; as well as a new problem that we call the Weighing Scale problem.

Contents

*University of Texas at Austin. These authors were supported in part by NSF award CCF-2225259.

†Cornell University. This author was supported by the Department of Defense (DoD) through the National Defense Science & Engineering Graduate (NDSEG) Fellowship Program.

‡Cornell University. This author was supported by NSF award CMMI-2307008.

1 Introduction

Many real-world settings involve algorithmic decision-making under incomplete information about the underlying input and future outcomes. In some cases, additional information can be acquired at a cost. A fundamental question arises: how should this costly information acquisition be incorporated into the algorithmic process?

Consider, for example, a biotechnology company running multiple drug discovery projects. These projects require upfront investment and entail a series of choices on which directions to pursue *before* their potential is known and realized. Other examples include an oil company's search for an optimal drilling site, a construction firm paying for multiple architectural designs before choosing one to pursue, and a manufacturer conducting market research before deciding which products to produce and in what quantity.

A classical model for costly information acquisition is the Pandora's Box problem, introduced by Weitzman [1979]. In this model, an algorithm faces an optimization problem over n stochastic alternatives, each concealed inside a closed box. The algorithm can open a box at a cost to observe the value of the alternative within, and can select an alternative only after its box has been opened. The goal is to minimize the total cost, which includes both the cost of opening the boxes and the values of the chosen alternatives.[1] The Pandora's Box model (henceforth, PB) is well understood and admits a simple optimal solution for a broad class of optimization problems. However, it is too simplistic to capture the complexities of real-world information acquisition: it assumes that each alternative can be revealed in only one way and that the algorithm either learns everything or nothing about it—there is no concept of partial information acquisition.

Recent research has explored extensions of PB that allow for multiple modes and stages of information acquisition.[2] These generalizations are typically NP-hard [Fu et al., 2022], necessitating the development of approximation algorithms. Beyhaghi and Cai [2024] provide a survey of recent advances in this domain. However, a significant limitation of prior work is that existing solutions are tailored to specific problem variants and do not extend to broader settings. For example, Aouad et al. [2020] address a two-stage inspection process, but their techniques do not generalize to three stages. Similarly, while previous results separately handle cases with partial inspection and optional inspection, it is unclear how to approach an instance that contains alternatives of both kinds.

In this paper, we develop a general framework for modeling *arbitrarily complex* protocols for information acquisition and designing approximation algorithms for them. Using this framework, we obtain improved approximations for several PB variants and develop algorithms for new models where non-trivial approximations were previously unknown.

A Framework for Modeling Information Acquisition. We model the process of costly information acquisition for each alternative $i \in [n]$ using a finite-horizon Markov decision process (MDP), denoted by $\mathcal{M}_i$. The current state of this MDP encapsulates all information the algorithm has about alternative i and determines the set of available actions for further exploration. Some actions lead to terminal states, where the alternative can be selected. For instance, both a standard optional-inspection PB and a three-stage inspection PB can be modeled as MDPs (see Figure 1). A key assumption in our framework is the independence of these MDPs: exploring one alternative does not affect the state of another.

This framework defines a broad family of optimization problems, which we call *Costly Information Combinatorial Selection* (CICS). In CICS, we are given n stochastic alternatives, each associated with an MDP $\mathcal{M}_1, \ldots, \mathcal{M}_n$. The goal is to select a subset of alternatives satisfying a given feasibility constraint; our work focuses on matroid feasibility constraints. The algorithm iteratively chooses an alternative i and takes an action in $\mathcal{M}_i$ to gather information. At any time, it can stop and select a feasible subset of alternatives that have reached terminal states. The objective function is the total expected value of the selected alternatives plus the cost of all actions taken.

The primary challenge in solving CICS arises from the interplay between two levels of decision-making: (i) *global decision-making*, which determines which alternative to explore at each step, and (ii) *local decision-*

[1] For maximization problems, the objective is to maximize the function value minus the cost of opening the boxes. In this section, we will focus our exposition on minimization problems although, as we show later in the paper, our framework and techniques extend to maximization problems as well.

[2] See, for example, Olszewski and Weber [2015], Kleinberg et al. [2016], Singla [2017], Doval [2018], Beyhaghi and Kleinberg [2019], Esfandiari et al. [2019], Gupta et al. [2019], Boodaghians et al. [2020], Aouad et al. [2020], Fu et al. [2022], Beyhaghi and Cai [2022], Berger et al. [2023], Bowers and Waggoner [2024a], Scully and Doval [2024].

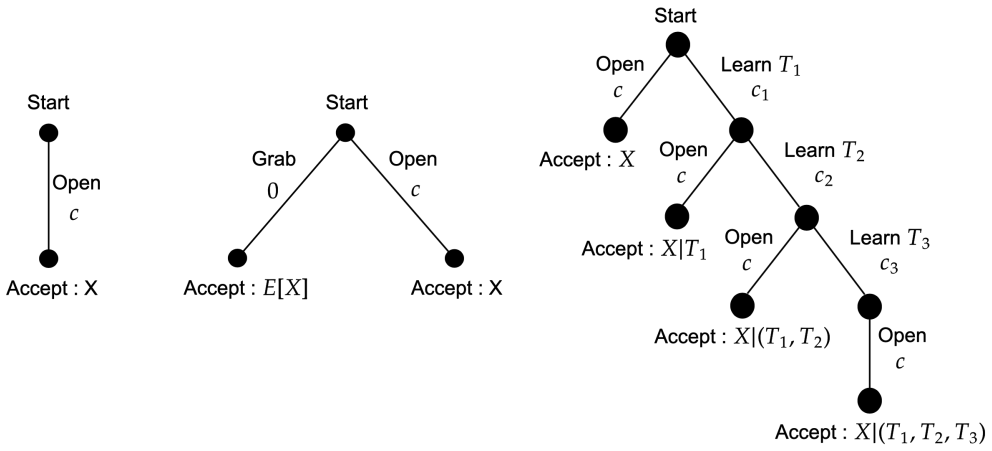


Figure 1: Three examples of a Pandora's Box style information acquisition protocol. (Left:) A classical PB is modeled as a Markov chain with a single costly action (opening the box) resulting in a random terminal state of value X. (Middle:) An optional-inspection PB modeled as an MDP; in addition to opening the box, there is a grab action that reveals no information and results in a single terminal state of deterministic value $\mathrm{E}[X]$. (Right:) A three-stage inspection protocol where the box can be opened directly, or it can be partially explored by sequentially revealing hidden variables T_i at a cost.

making, which dictates the costly actions taken within each alternative's MDP. These decisions are typically adaptive, responding to the information that was acquired in previous steps. Since CICS generalizes MDPs, we have no hope of obtaining a general purpose solution to the problem. Instead, what we aim for is a decomposition of the problem into smaller pieces that can each be solved independently and composed together into a global solution. We obtain such a decomposition by analyzing a class of algorithms called *committing policies*.

Committing Policies and the Commitment Gap. A *committing policy* for CICS is a policy that, in essence, predetermines all local decisions before making any global decisions. In particular, it preselects an action for each state in every MDP $\mathcal{M}_i$, effectively reducing the MDPs to Markov chains. This is valuable because for the class of feasibility constraints we are interested in, the special case of CICS where each MDP $\mathcal{M}_i$ is effectively a Markov chain, i.e. has just one action in every state, admits a simple optimal policy [Dumitriu et al., 2003, Gupta et al., 2019]. By fixing local decisions in advance, we significantly reduce the complexity of global decision-making. This prompts two questions:

- What is the best approximation ratio achievable by committing policies for a given class of CICS problems? We call this ratio the *commitment gap*.
- Can the local decision commitments be made in a *composable* way? That is, can the commitments for each MDP $\mathcal{M}_i$ be determined using only the characteristics of $\mathcal{M}_i$, independent of other MDPs?

These questions have been posed and investigated for optional-inspection PB [Beyhaghi and Kleinberg, 2019, Scully and Doval, 2024], but little is known beyond that setting. See Section 2 for further discussion of prior work.

In this paper, we address both questions by developing a set of *local* sufficient conditions for bounding the commitment gap in general CICS. By "local", we mean that our conditions depend only on the properties of each individual MDP $\mathcal{M}_i$, without reference to the other MDPs or the feasibility constraint. Furthermore, our conditions yield policies that determine their commitments locally (or, in one case, with a simple global algorithm after local pre-processing). We introduce two key conceptual contributions that enable bounds on the commitment gap: a method of bounding the optimum, and notions of approximation ratio that can be checked locally for each MDP.

Bounding the Optimum. Our first conceptual contribution is a lower bound on the cost of the (unrestricted, non-committing) optimum. This bound is analogous to the Whittle integral [Whittle, 1980] for infinite-horizon

discounted-reward MDPs. While prior proofs of the Whittle integral exist for the discounted-reward setting [Brown and Smith, 2013, Hadfield-Menell and Russell, 2015], we present an algorithmic proof for the finite-horizon case, which is simpler and provides additional insight.

Importantly, our approach extends the notion of *surrogate costs* from PB literature to general MDPs and connects it to the Whittle integral, offering valuable algorithmic insights into bounding the commitment gap. Surrogate costs were first defined for classical PB by [Kleinberg et al., 2016] and [Singla, 2017], and can be extended easily to arbitrary Markov chains [Weber, 1992, Dumitriu et al., 2003, Gupta et al., 2019]. However, the extension to MDPs is highly non-trivial and is a novel contribution of our work. We describe the details in Section 1.1.

Local Approximation. Our second contribution is the development of a *local approximation* technique, which expresses the global approximation factor of a committing policy $(\pi_1, \ldots, \pi_n)$ in terms of the local properties of each constituent policy π_i relative to its corresponding MDP $\mathcal{M}_i$. This concept was first introduced by Scully and Doval [2024] for the optional-inspection PB; we generalize it to arbitrary MDPs.

Crucially, local approximations compose: if each π_i achieves an α-local approximation for $\mathcal{M}_i$, then the committing policy $(\pi_1, \ldots, \pi_n)$ achieves an α-global approximation. This observation answers, for example, our aforementioned question of how to compose partial-inspection Pandora's Boxes with optional-inspection ones within a single instance. Furthermore, our global approximation guarantee extends to any combinatorial setting that admits a "greedy-style" algorithm via techniques similar to [Singla, 2017, Gupta et al., 2019].

Although local approximation can be defined directly with respect to the Whittle integral, as in [Scully and Doval, 2024], proving guarantees using this definition appears very challenging. One of our main contributions is to connect this concept to surrogate costs in a manner that enables simpler analysis. We develop multiple approaches for bounding local approximation factors that are suitable for different contexts, and demonstrate their use through the applications discussed in Section 1.1.

Example Application: Shortest Paths. To illustrate the power of our framework, consider the *Pandora's Shortest Path* problem introduced by Singla [2017]: the edge weights of a given graph correspond to independent Pandora's boxes and the objective is to accept a set of (open) boxes that form a path between two given vertices s and t, while minimizing the total cost of opening boxes plus the cost of the path. As path constraints do not admit frugal algorithms that enable the generalization of Weitzman's result to combinatorial settings, this problem is notoriously challenging and no approximation results are known.

When the underlying graph is a union of disjoint s-t paths, this problem becomes an instantiation of CICS: each path corresponds to an alternative, whose MDP encapsulates the different orders in which the constituent "edge boxes" can be opened. A committing policy, in this case, pre-defines a protocol for the order of inspecting the edges within a path, that can adapt only to the instantiations of the edges from the same path. We show that the commitment gap of this special case of Pandora's Shortest Paths is at most 2.

1.1 Our technical contributions and their relationship to prior work.

We now describe our technical contributions in more detail in the context of closely related prior work. A thorough discussion of other related work is presented in Section 2.

Bandit superprocesses and the Whittle integral. CICS is closely related to a class of sequential decision-making problems called bandit superprocesses (BSPs), introduced by [Nash, 1973]. A BSP is composed from multiple independent MDPs; at every step, the algorithm advances one of the MDPs, receives a (discounted) reward, and the state of the underlying MDP evolves stochastically. A primary difference between the two settings is that a CICS algorithm must terminate by selecting a feasible set of alternatives, whereas a BSP can potentially have an infinite horizon. Moreover, since the states of the MDPs in a CICS model information acquisition, we assume that each constituent MDP is acyclic; our techniques rely on this acyclicity. In contrast, BSPs can be defined over arbitrary MDPs.

In the special case where the MDPs are Markov chains, BSP reduces to the extensively-studied Multi-Armed Bandit (MAB) problem. MAB admits an astonishingly simple optimal solution [Gittins, 1979]: every state in each constituent Markov chain is assigned an index independent of the other chains; at every step, the Markov chain with the maximum index in its current state is advanced. [Dumitriu et al., 2003] show that this "indexability"

extends to the finite-horizon setting as well, and [Gupta et al., 2019] further extend it to combinatorial settings. We restate this result with a new proof as Theorem 4.4. [Weitzman, 1979]'s algorithm for classical PB is essentially a special case of this result.

General BSPs, in contrast, are not indexable [Glazebrook, 1982].[3] Nevertheless, the local structure of the problem provides insight into the optimal cost: given an MDP $\mathcal{M}$, [Whittle, 1980] considers a local problem $(\mathcal{M}, y)$, where at every step the algorithm can either advance $\mathcal{M}$ or terminate with a cost of y. The values of these local games, formalized as "optimality curves" $f_{\mathcal{M}}(y)$, one for each constituent MDP in the BSP, can be combined to obtain the Whittle integral, a lower bound on the global optimum [Whittle, 1980, Brown and Smith, 2013].

Our work provides a new interpretation for Whittle's optimality curves by connecting them with a mapping from trajectories in the MDP to "surrogate" costs. The surrogate costs essentially allow an algorithm to amortize the costs of actions and pay (a part of) them only when the MDP terminates. We develop a recursive water filling algorithm for cost amortization that ensures that good, i.e. low cost or high reward, terminal states are responsible for paying most of the cost share. A crucial property we achieve is the *independence* of the distribution of surrogate costs from the actions chosen by an algorithm in the MDP in the local or global game. In particular, the costs capture the local structure of the MDP without limiting in any way how an algorithm "solves" the MDP. This independence allows us to compose them into a global bound for the optimal (adaptive, non-committing) policy.

Our approach is heavily influenced by [Kleinberg et al., 2016] and [Singla, 2017]'s amortization viewpoint on Weitzman's index for classical PB, as well as its extension to Markov chains. However, we emphasize that the extension to MDPs while ensuring action independence is an important and novel contribution of our work that enables our approximation bounds.

Commitment gap and local approximation. We follow the approach of [Scully and Doval, 2024] to establish a bound on the commitment gap by quantifying the performance of each commitment π_i relative to the optimal policy for MDP $\mathcal{M}_i$ in the local game $(\mathcal{M}_i, y)$. Such an approach helpfully disentangles global and local decision making.

Local guarantees for $(\mathcal{M}_i, y)$ need to be established with care, however. [Whittle, 1980] showed that if every $\mathcal{M}_i$ admits a commitment π_i that is optimal for the local game $(\mathcal{M}_i, y)$ regardless of the value y, then the tuple $\mathcal{P} = (\pi_1, \ldots, \pi_n)$ is globally optimal. In other words, unambiguous local optimality composes into global optimality. However, the reverse doesn't hold: the globally optimal policy might take actions that are unambiguously *suboptimal* in the local game! (See Appendix C.) Likewise, simply relating the optimal cost $f_{\mathcal{M}_i}(y)$ to the cost of the commitment $f_{\mathcal{M}_i^{\pi_i}}(y)$ at every y is not sufficient, as we demonstrate in Appendix C.

Following [Scully and Doval, 2024] we say that committing policy π_i α-locally approximates $\mathcal{M}_i$ if for all y we have $f_{\mathcal{M}_i^{\pi_i}}(\alpha y) \leq \alpha \cdot f_{\mathcal{M}_i}(y)$. Local approximations defined in this manner compose into global guarantees on the commitment gap. We use this approach to obtain new approximations for PB with partial inspections (defined below).

Unfortunately, however, local approximations can be very challenging to prove for general MDPs. We develop a new (but weaker) local condition based on surrogate costs that we call *pointwise approximation*. This allows us to analyze more complicated MDPs with multiple rounds of actions for which optimality curves are tricky to establish.

Finally, the commitment gap can sometimes be strictly smaller than what can be established through local approximation. We investigate this issue in the context of PB with optional inspection and show that improved results can be obtained through a *semi-local* argument.

Applications of our framework. We instantiate our framework to obtain new bounds on the commitment gap of four PB extensions. The first three variants are new problems studied in the minimization setting, and we obtain the following results:

- **PB with Partial Inspection**. In addition to opening the box, the algorithm can choose to "peek" into it at a cheaper cost and learn its value; boxes must still be opened before being selected. We prove that the

[3]This is because MAB only involves global decision-making and does not require any local decision-making as there is only one way to advance the chosen Markov chain.

commitment gap is at most $\sqrt{2}$.

- **Additive PB**. The value of the box is given by the sum of independent random variables; each component must be separately inspected at a cost and the box can be selected only after all components have been inspected. This models the shortest paths problem over disjoint paths described earlier. We prove that the commitment gap is at most 2.
- **Weighing Scale Problem**. The box cannot be opened, but can be compared at a cost against a threshold t (chosen by the algorithm); the comparison reveals whether $X_i \geq t$ or not. Any sequence and number of comparisons are allowed. We obtain a logarithmic bound (in key parameters of the problem) on the commitment gap.

All these results apply to any set of matroid feasibility constraints, and can be seamlessly extended to any set of constraints that admit a frugal algorithm, as defined by [Singla, 2017]. Furthermore, they immediately extend to settings that consider combinations of these variants: an instance consisting of some (normal) boxes, some partial inspection boxes and some additive boxes under matroid feasibility constraints still has a commitment gap of at most 2.

We also study the commitment gap in a maximization setting in the context of **PB with Optional Inspection**. This problem has been extensively studied in the the single-selection setting [Doval, 2018, Guha et al., 2008, Beyhaghi and Kleinberg, 2019, Beyhaghi and Cai, 2024], culminating in the discovery of a PTAS [Fu et al., 2022, Beyhaghi and Cai, 2022]. However, aside from the simultaneous work described below, the maximization version of the problem has not been studied in the combinatorial setting, with Scully and Doval [2024] covering only the minimization setting. While the PB local approximation result of Scully and Doval [2024] can be extended to maximization without too much trouble, it yields a worst-case local approximation ratio of only 0.5, which can also be achieved by a trivial single-coin-flip algorithm [Beyhaghi and Kleinberg, 2019]. To remedy this, we introduce an extension of local approximation, which we call *semilocal approximation*, that introduces just a small amount of dependence between the randomized commitments for each box. We use semilocal approximation to construct a committing policy that is a 0.58-approximation under matroid feasibility constraints. While this worst-case ratio is not state-of-the-art,[4] we include our construction because (i) we obtain instance-dependent bounds on the commitment gap that are better in some regimes, (ii) our construction is simpler and more explicit than other known approaches that beat the 0.5-approximation barrier, and (iii) semilocal approximation may prove useful for other PB variants in the future.

Simultaneous and Subsequent Work. In simultaneous and independent work, [Bowers et al., 2025] also study CICS in the maximization setting. Rather than approaching the problem through local approximation, they consider an ex-ante relaxation of the optimal policy that allows them to efficiently obtain a non-committing approximation policy in settings where the underlying combinatorial constraint admits a prophet inequality; in particular, they obtain a 0.5-approximation via an adaptive policy under matroid constraints. This approach was subsequently further explored and significantly generalized by Chawla et al. [2025], establishing bounds on the commitment gap. A key difference between these works and our framework is that (i) our results and local approximation also apply to the minimization setting, and (ii) the bounds we obtain on the commitment gap also depend on the structure of the underlying MDPs, allowing for improved approximations in some instances.

1.2 Outline of the rest of the paper. A detailed discussion of related work is presented in Section 2. In Section 3, we provide formal definitions for CICS and the commitment gap. Our amortization framework and bound on the optimal policy for the minimization setting are presented in Section 4; corresponding results for the maximization setting are established in Appendix A. The local approximation framework and its composition to bounds for the commitment gap are presented in Section 5. Subsequent sections are dedicated to bounding

[4]While no better approximation algorithms were published prior to this work, we discovered during the revision process that Beyhaghi and Kleinberg [2019]'s existential bound of $(1 - 1/e)$ on the commitment gap can be matched efficiently up to arbitrary precision. Informally, their approach is based on a reduction to nonadaptive stochastic submodular maximization, maximizing the weighted rank function of the matroid constraint subject to another partition matroid. While achieving exactly a $1 - 1/e$ ratio in the submodular maximization step is nonconstructive, it can be made polynomial-time constructive at the cost of arbitrarily small error [Asadpour and Nazerzadeh, 2016].

the commitment gap for different instantiations of CICS: PB with Partial Inspection (Section 6), Additive PB (Section 7), Weighing Scale Problem (Section 8) and Optional Inspection PB (Section 9).

2 Related work

We have already discussed the relationship of CICS to **bandit superprocesses** and Markovian multi-armed bandit. There is little work on BSPs that do not satisy Whittle's "unambiguous local optimality" condition, outside of the variants of Pandora's Box. The only such work we are aware of is by [Ke and Villas-Boas, 2019], who look at a problem with two specific symmetric MDPs and an arbitrary outside option, and develop an adaptive (but complicated) exact algorithm for this problem. We refer the reader to the survey by [Hadfield-Menell and Russell, 2015] for further discussion of BSPs.

Combinatorial variants of Pandora's Box were first studied by [Singla, 2017] who showed that for feasibility constraints admitting greedy-style or "frugal" approximation algorithms, an extension of Weitzman's indexing algorithm provides the same approximation factor. This was further extended to CICS over Markov chains by [Gupta et al., 2019].

The idea of augmenting the inspection process of a Pandora's box in order to explore more interesting decision-making settings has been a very active line of research over the last years. Arguably, the most studied variant is **optional inspection**. Most literature focuses on the single-item selection, as opposed to combinatorial variants. Study of it was initiated by Guha et al. [2008], who give a 4/5-approximation for the maximization setting; and Doval [2018], who characterized the solution to the single-box problem and proved certain conditions under which the Gittins policy remains optimal for single-item selection in the maximization setting (though the results naturally extend to the minimization setting). Beyond this results are separated by whether they are for the minimization or maximization setting. For the minimization setting, Scully and Doval [2024] proved a composition theorem for a special case of local approximation (Definition 5.1) and used it to construct a committing policy with a 4/3-approximation guarantee. In fact, their result extends beyond the single-item selection setting to the combinatorial Pandora's box setting studied by Singla [2017]. In the maximization setting, Beyhaghi and Kleinberg [2019] and Guha et al. [2008] give approximation guarantees for committing policies. Furthermore, Fu et al. [2022] and Beyhaghi and Cai [2022] introduced polynomial time approximation schemes that for any $\varepsilon > 0$ compute a policy that is at least a $(1-\varepsilon)$-approximation. However, all of these results are for the single-item selection setting. For matroid feasibility constraints, Beyhaghi and Kleinberg [2019] observe that the commitment gap is at least $(1-1/e)$ and give an efficient policy that achieves a 0.5 approximation; as already mentioned, their ideas can be pushed further to efficiently obtain a $(1-1/e-\epsilon)$ approximation for any $\epsilon > 0$.

The **partial inspection** variant of Pandora's Box has also primarily been studied in the context of maximization. [Aouad et al., 2020] provide a $(1-1/e)$-approximation via a committing policy, and show that, in fact, *any* committing algorithm or its negation (flipping which box should be partially opened versus fully opened) admits a (1/2)-approximation to the optimal utility. We note that this already highlights a significant difference between the minimization and maximization settings for this variant. Whereas for maximization one can essentially flip a coin to decide which of the two actions to commit to, obtaining the optimal's utility for the committed action and non-negative utility for the action that was not selected, the same approach cannot be applied for minimization as the cost suffered by a bad flip could result to arbitrarily bad approximations. Beyhaghi [2019] introduces a more general inspection model, where the searcher has k different methods for inspecting each box, and can select at most one of them. They provide a $(1-1/e)$-approximation that applies to k-element selection, but is computationally inefficient when k is large. To our knowledge, partial inspection has not been studied in the context of minimization.

The **additive PB** model has been studied previously [Bowers and Waggoner, 2024a,b] in the context of maximization for $k = 2$ random variables per alternative with matching constraints. In this case, each alternative admits only two commitments: opening one of the two random variables followed by the other. The authors show via a global argument that picking one of the two commitments randomly independently for each alternative yields a bound of 2 on the commitment gap (a further factor of 2 is lost due to the matching constraint). For boxes with $k > 2$ components each, applying the same approach implies a commitment gap of $O(k!)$. We leave as an open question whether a better bound can be obtained using our techniques. The minimization variant of this problem has not been explored in prior research.

It is noteworthy that a common approach in the literature to bounding the **commitment gap** in maximization settings involves reducing the problem to finding the optimal committing policy for a stochastic submodular

maximization instance and applying the bound of [Asadpour and Nazerzadeh, 2016] on the adaptivity gap for such problems [Beyhaghi and Kleinberg, 2019, Aouad et al., 2020, Beyhaghi, 2019]. This method is inherently global, as it requires efficient optimization over the entire instance once mapped to a stochastic submodular maximization framework. Moreover, the optimization scales with the number of committing policies and does not extend beyond the single-selection case.

Other extensions of Pandora's Box include settings with combinatorial rewards [Olszewski and Weber, 2015]; combinatorial costs [Berger et al., 2023]; correlated values [Chawla et al., 2019, 2021, Gergatsouli and Tzamos, 2024]; or constraints on the order of inspection [Esfandiari et al., 2019, Boodaghians et al., 2020, Bowers and Waggoner, 2024a]. We refer the reader to the survey by [Beyhaghi and Cai, 2024] for further discussion of Pandora's Box.

The **weighing scale** problem has not been studied previously, although similar settings where the algorithm is provided with a *budget* on the total number of weighings it can perform have been considered. In particular, Hoefer et al. [2024] study a setting, where each alternative can be weighed against the *median* (or any other fixed quantile) of its distribution (conditioned on any past weighings). Hoefer and Schewior [2023] also consider the setting where the alternatives are identically distributed and the decision-maker is allowed to weigh them against any threshold of their choosing (much like the weighing scale problem). In both of these works, using the weighing scale is free; the algorithm is provided with a budget on the number of weighings it can perform; the objective is to find the best alternative subject to the budget; and constant factor approximations are obtained.

3 Definitions

We begin by defining costly information acquisition for a random variable as a Markov decision process (MDP). In a Costly Information MDP, states represent the information the algorithm has gained about the corresponding random variable. Accordingly, we associate each state with the conditional value distribution it represents. Formally:

Definition 3.1. (Costly Information MDP) *A Costly Information MDP for a random variable X is a tuple $\mathcal{M}_X = (S, \sigma, A, c, \mathcal{D}, V, T)$, where S is a set of states, $\sigma \in S$ is the starting state, $A(\cdot)$ maps states to sets of actions, $c(\cdot)$ is a cost function mapping actions to costs, and $\mathcal{D}$ is a transition matrix. For each pair of states $s, s' \in S$ and each action $a \in A(s)$, $\mathcal{D}(s, a, s') \in [0, 1]$ specifies the probability of transitioning to s' upon taking action a in state s; naturally, $\sum_{s'} \mathcal{D}(s, a, s') = 1$ for all states $s \in S$ and actions $a \in A(s)$.*

$V(\cdot)$ is a function mapping states to distributions over values. $V(\sigma)$ is the (unconditional) distribution of X and $V(s)$ is the posterior distribution of X conditioned on being in state $s \in S$. As such, V satisfies the rules of conditional probability: for all states $s \in S$ and all actions $a \in A(s)$, we have $V(s) = \sum_{s'} \mathcal{D}(s, a, s')V(s')$. We also write $v(s) := \mathrm{E}[V(s)]$. Finally, $T \subseteq S$ is the set of terminal states. Terminal states have only one action available, called the "accept" action. This accept action comes at no cost; results in a value of $v(s)$ at terminal state $s \in T$; and terminates the MDP. In the special case where there is only one action (accept or advance) available at every state $s \in S$, we call $\mathcal{M}_X$ a Costly Information Markov chain. When clear from the context, we will drop the subscript X.

Assumptions. Note that each costly information MDP $\mathcal{M}_X$ is inherently acyclic (i.e. the underlying graph is a DAG), as each costly action further refines the algorithm's knowledge of the underlying random variable. For simplicity of exposition and without loss of generality, we assume that the state spaces, action sets, and the support of the random variable X are finite, although our framework and results extend seamlessly to continuous settings. We will also assume a bounded horizon. In particular, there exists a constant H such that any state reached after a sequence of H actions is terminal.

Definition 3.2. (Costly Information Combinatorial Selection) *The Costly Information Combinatorial Selection problem (henceforth, CICS) is defined over a ground set of n nonnegative random variables, $X_1, X_2, \cdots, X_n$; a costly information MDP $\mathcal{M}_i := \mathcal{M}_{X_i} = (S_i, \sigma_i, A_i, c_i, \mathcal{D}_i, V_i, T_i)$ for each variable X_i; and a feasibility constraint $\mathcal{F} \subseteq 2^{[n]}$. The constraint $\mathcal{F}$ corresponds to an upwards closed set in the minimization version (henceforth, min-CICS), and to a downwards closed set in the maximization version (henceforth, max-CICS). We use $\mathbb{M} := (\mathcal{M}_1, \cdots, \mathcal{M}_n)$ to denote the set of MDPs and $\mathbb{I} = (\mathbb{M}, \mathcal{F})$ to denote the CICS instance.*

General Policies for CICS. A policy (a.k.a. algorithm) for CICS proceeds as follows. Let $S \subseteq [n]$ denote the set of indices of all terminated MDPs. Let s_i denote the current state of the MDP $\mathcal{M}_i$ at any point of time during the process. Initially, $S = \emptyset$ and $s_i = \sigma_i$ for all $i \in [n]$. The algorithm chooses at every step an index $i \in [n] \setminus S$ corresponding to a non-terminated MDP $\mathcal{M}_i$ and an action $a_i \in A_i(s_i)$. It then follows the action at a cost of $c_i(a_i)$. If a_i is the accept action (i.e. $s_i \in T_i$ is a terminal state), it adds i to S and collects the value $v_i(s_i)$. Otherwise, it updates the state of $\mathcal{M}_i$ to a new state drawn from the distribution $\mathcal{D}(s_i, a_i, \cdot)$, while the states of all other MDPs $\mathcal{M}_{i'}$ with $i' \neq i$ remain unchanged. Observe that the algorithm can make both of these choices – the index of the MDP to move in and the action to take in that MDP – adaptively, depending on the evolution of all n MDPs in previous steps.

- For min-CICS, the algorithm terminates as soon as $S \in \mathcal{F}$. The objective of min-CICS is to find an algorithm with *minimum total cost*, defined as the expectation (over the randomness of the algorithm and the underlying processes) of the total cost of all actions undertaken by the algorithm until termination *plus* the values accrued from accept actions.
- For max-CICS, the algorithm needs to ensure feasibility by selecting at every step indices $i \in [n] \setminus S$ such that $S \cup \{i\} \in \mathcal{F}$. The objective of max-CICS is to find an algorithm with *maximum utility*, defined as the expectation (over the randomness of the algorithm and the underlying processes) of the total value accrued from accept actions *minus* the total cost of all actions undertaken by the algorithm until termination.

We use $\mathrm{OPT}(\mathbb{I})$ to denote the cost/utility of the optimal policy for the CICS instance $\mathbb{I} = (\mathbb{M}, \mathcal{F})$.

Committing Policies for CICS. A commitment π_i for the MDP $\mathcal{M}_i$ is a mapping from every state $s_i \in S_i$ of $\mathcal{M}_i$ to a distribution $\pi_i(s_i)$ over the actions in $A_i(s_i)$. A committing policy is then characterized by a tuple of commitments $\mathcal{P} = (\pi_1, \cdots, \pi_n)$; at each step, the policy selects an MDP $\mathcal{M}_i$ at current state s_i and takes an action that is sampled from $\pi_i(s_i)$. Formally, for $i \in [n]$, let $\mathcal{M}_i^{\pi_i}$ denote the Markov chain resulting from applying the commitment π_i to MDP $\mathcal{M}_i$. Then any policy for the instance $\mathbb{I}_{|\mathcal{P}} := (\mathcal{M}_1^{\pi_1}, \cdots, \mathcal{M}_n^{\pi_n}, \mathcal{F})$ is a committing policy for $\mathbb{I}$ consistent with the tuple $\mathcal{P}$. Note that while committing policies must make all local decisions as dictated by $\mathcal{P}$, the index of the MDP to move in can be selected adaptively: committing policies are therefore adaptive algorithms.

We use $\mathcal{C}(\mathcal{M}_i)$ to denote the set of all possible commitments π_i for the MDP $\mathcal{M}_i$ and $\mathcal{C}(\mathbb{I}) = \prod_i \mathcal{C}(\mathcal{M}_i)$ to denote the set of all possible tuples of commitments $\mathcal{P}$ for the CICS instance $\mathbb{I}$. The commitment gap quantifies the performance loss of committing policies relative to the optimal policy.

Definition 3.3. (Commitment Gap) *The commitment gap for any min-CICS (or max-CICS) instance* $\mathbb{I} = (\mathbb{M}, \mathcal{F})$ *is defined as*

$$\mathrm{CG}(\mathbb{I}) := \min_{\mathcal{P} \in \mathcal{C}(\mathbb{I})} \frac{\mathrm{OPT}(\mathbb{I}_{|\mathcal{P}})}{\mathrm{OPT}(\mathbb{I})} \geq 1 \quad \Big(\textit{or} \;\; \mathrm{CG}(\mathbb{I}) := \max_{\mathcal{P} \in \mathcal{C}(\mathbb{I})} \frac{\mathrm{OPT}(\mathbb{I}_{|\mathcal{P}})}{\mathrm{OPT}(\mathbb{I})} \leq 1 \Big).$$

Matroid-min-CICS. From now on, we focus primarily on matroid feasibility constraints: each MDP corresponds to an element of some known matroid $\mathrm{M} = ([n], \mathcal{I})$. For min-CICS, $\mathcal{F}$ is the collection of all sets that contain a basis of M; for max-CICS, $\mathcal{F}$ contains all the independent sets of M. Observe that single-element selection, where the algorithm's goal is to accept one MDP, is a special case of matroid selection. Following the work of Singla [2017], our results will apply to any feasibility constraint that admits an efficient *frugal approximation algorithm*. We describe this extension in Appendix B. For conciseness, in the main body of the paper, we state our results for the minimization setting (matroid-min-CICS); the analogous framework and results for the maximization setting are discussed in Appendix A.

4 An Amortization Framework

In this section, we develop a novel cost amortization technique that will allow us to lower bound the cost of the optimal adaptive algorithm for matroid-min-CICS. We first establish our amortization for the special case of Markov chains (Section 4.1). We then connect our amortization framework to the notion of optimality curves (Section 4.2) and use this connection to extend our approach to general MDPs (Section 4.3). For ease of presentation, we present all our proofs in Section 4.4.

4.1 Amortized Surrogate Costs for Markov Chains. In this section we consider instances $\mathbb{I} = (\mathbb{M}, \mathcal{F})$ of the matroid-min-CICS where every MDP $\mathcal{M}_i$ is a Markov chain. We begin with some notation. For a Markov chain $\mathcal{M} = (S, \sigma, A, c, \mathcal{D}, V, T)$ and a state $s \in S$, we use $c(s)$ to denote the cost of the unique action in $A(s)$. A *trajectory* τ in this Markov chain is a sequence of states that results from advancing the chain until it terminates. For a state $s \in S$, let $\Upsilon(s)$ denote the set of all possible trajectories starting in s and let $\Upsilon := \Upsilon(\sigma)$. For state $s \in S$ and trajectory $\tau \in \Upsilon(s)$, let $p(\tau)$ denote the probability that a random trajectory starting in the initial state σ visits the state s and then continues along τ until terminating. Let $p(s) := \sum_{\tau \in \Upsilon(s)} p(\tau)$ denote the probability of visiting s. For $\tau \in \Upsilon$ and $s \in \tau$, let τ_s denote the suffix of τ starting with s. Finally, let $v(\tau)$ be the value $v(t)$ of the terminal state t that τ ends in.

We will now define an amortization of action costs in a Markov chain over the resulting (random) trajectories that follow. The intention is that in the amortized setting, the algorithm does not incur any action costs when the action is taken but pays the amortized cost of the instantiated trajectory *only if* it follows the trajectory all the way to termination.

Definition 4.1. (Cost Amortization) *A cost amortization of a Markov chain $\mathcal{M} = (S, \sigma, A, c, \mathcal{D}, V, T)$ is a non-negative vector $b = \{b_{s\tau}\}_{s \in S, \tau \in \Upsilon(s)}$ with the property that $\sum_{\tau \in \Upsilon(s)} p(\tau) b_{s\tau} = p(s)c(s)$ for all states $s \in S$. Based on this amortization, we define:*

- *The amortized cost of a trajectory $\tau \in \Upsilon$ as $\rho_b(\tau) := v(\tau) + \sum_{s \in \tau} b_{s\tau_s}$.*
- *The surrogate cost of the Markov chain $\mathcal{M}$ as the random variable $\rho_{\mathcal{M},b}$ that takes on value $\rho_b(\tau)$ for $\tau \in \Upsilon$ with probability $p(\tau)$.*
- *The index of a state $s \in S$ of the Markov chain $\mathcal{M}$ as $I_{\mathcal{M},b}(s) := \min_{\tau \in \Upsilon : s \in \tau} \rho_b(\tau)$.*

Note that the "cost shares" $b_{s\tau}$ distributed by a state s towards its resulting trajectories fully cover the cost of s's action and no more. However, in the amortized setting, these cost shares are paid only when the trajectory terminates, and go unpaid if the algorithm stops advancing the Markov chain midway through the trajectory. We therefore obtain the following lemma.

Lemma 4.2. *Consider any instance $\mathbb{I} = (\mathbb{M}, \mathcal{F})$ of matroid-min-CICS over Markov chains and let b_i be any cost amortization of $\mathcal{M}_i$ with surrogate cost $\rho_i := \rho_{\mathcal{M}_i, b_i}$ for all $i \in [n]$. Then,* $\mathrm{OPT}(\mathbb{I}) \geq \mathrm{E}\left[\min_{S \in \mathcal{F}} \sum_{i \in S} \rho_i\right]$.

We will now exhibit a specific cost amortization and a corresponding algorithm that achieves the lower bound of Lemma 4.2, proving optimality. The **water filling cost amortization** is described algorithmically in a bottom-up fashion. We consider the states of the chain in reverse topological order, starting from the terminals up towards the root σ. Trajectories $\tau \in \Upsilon(t)$ for terminal states $t \in T$ are singletons and do not carry cost shares. Consider a state s such that the cost shares for all states reachable from it have been determined. The state distributes its total cost $c(s)$ across its downstream trajectories $\tau \in \Upsilon(s)$, starting from those with the **lowest current total cost**, until the equation $\sum_{\tau \in \Upsilon(s)} p(\tau) b_{s\tau} = p(s)c(s)$ is satisfied. Formally, let $\tau' = \tau \setminus s$ be the sequence of states in τ following s; its current downstream cost is $\rho(\tau') = v(\tau') + \sum_{s' \in \tau'} b_{s'\tau'_{s'}}$. We find the lowest water level g such that the cost shares $b_{s\tau} := (g - \rho(\tau \setminus s))^+$ satisfy the cost equation above. The new total cost of τ becomes $\rho(\tau) = b_{s\tau} + \rho(\tau \setminus s) = \max\{g, \rho(\tau \setminus s)\}$. We continue in this manner until σ's cost is amortized.

We use $W^*_{\mathcal{M}}$ to denote the water filling surrogate cost of a Markov chain $\mathcal{M}$, and $I^*_{\mathcal{M}}(s)$ to denote the water filling index of a state s in $\mathcal{M}$. Finally, we can describe an index-based policy based on the water filling amortization.

Definition 4.3. (Water Filling Index Policy) *The water filling index policy for an instance $\mathbb{I} = (\mathbb{M}, \mathcal{F})$ of matroid-min-CICS over Markov chains selects at every step the Markov chain $i^* = \mathrm{argmin}_{i \in \mathcal{F}_S} I^*_i(s_i)$, where s_i is the current state of Markov chain $\mathcal{M}_i$; S is the set of terminated (selected) Markov chains; and $\mathcal{F}_S = \{i : \mathit{rank}(S \cup \{i\}) > \mathit{rank}(S)\}$.*

Note that, by definition, the indices of states encountered on any trajectory weakly increase with each next action. The water filling amortization ensures that if $b_{s\tau} > 0$ for some state $s \in S$ and $\tau \in \Upsilon(s)$, then the index of all states in the downstream trajectory τ stays *equal to* the index of s. Since the policy described above always advances the Markov chain of minimum index, this implies that for any action taken by the policy, any downstream

trajectory that "owes" a non-zero cost share to that action will terminate with acceptance if instantiated. In other words, all the amortized cost shares are paid in expectation, establishing the following optimality result:

THEOREM 4.4. *For any matroid-min-CICS instance* $\mathbb{I} = (\mathbb{M}, \mathcal{F})$ *over Markov chains, the expected cost of the water filling index policy is equal to* $\mathrm{E}\left[\min_{S\in\mathcal{F}} \sum_{i\in S} W^*_{\mathcal{M}_i}\right]$. *The policy is therefore optimal for* $\mathbb{I}$.

To conclude this section, we will illustrate water filling through an example.

EXAMPLE 4.5. *Consider an MDP* $\mathcal{M}$ *whose starting state is given a choice between following one of two Markov chains* $\mathcal{M}_1$ *and* $\mathcal{M}_2$. *Each Markov chain has a single costly action and results in a distribution over two distinct states, as illustrated in Figure 2. The action costs, terminal values, and instantiation probabilities are indicated in the figure. For each chain, the water filling amortization computes a unique index* g *(here,* $g_1 = 2$ *and* $g_2 = 1$*), corresponding to the value for which the highlighted area equals the cost of the amortized action. Each trajectory leads to a unique terminal node, so we can associate cost shares with terminals. For each terminal* t, *the corresponding cost share is given by* $(g - v(t))^+$; *here,* $b_{11} = 4/3$, $b_{21} = 1/2$ *and* $b_{12} = b_{22} = 0$. *Therefore, the surrogate cost* $W^*_{\mathcal{M}_1}$ *is* 2 *with probability* (3/4) *and* 4 *with probability* (1/4). *Likewise,* $W^*_{\mathcal{M}_2}$ *is* 1 *with probability* (1/4) *and* 3 *with probability* (3/4).

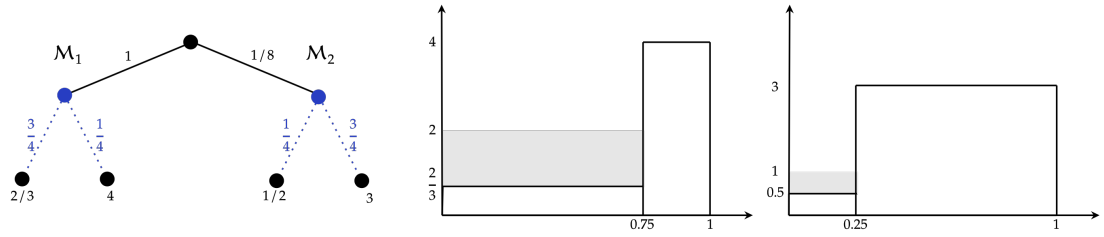


Figure 2: The water filling amortization for Markov chains $\mathcal{M}_1$ (left branch) and $\mathcal{M}_2$ (right branch). The solid black edges denote costly actions (of costs 1 and 1/8) and the dashed blue edges denote random transitions to terminal states upon taking these actions, with the corresponding probabilities of transition. Each terminal state has a corresponding value, depicted in the bottom line. The graphs on the right display the water-filling operation for each chain with the x-axis representing probabilities of instantiation of the respective trajectories and the y-axis representing the terminal values for these trajectories. The area of the gray region equals the action cost and its height is the water level that determines cost shares. By definition of water-filling, cost shares are first transferred to the trajectories of minimum value.

4.2 Local Games and Optimality Curves. We now connect the water filling amortization described above to the notion of optimality curves for MDPs, defined in the work of Whittle [1980] and its follow ups. We first define a "local game" for an MDP $\mathcal{M}$.

DEFINITION 4.6. (LOCAL GAME AND OPTIMALITY CURVE) *The local game* $(\mathcal{M}, y)$ *is a single-selection min-CICS with two MDPs, one of which is the MDP* $\mathcal{M}$. *The second MDP, a.k.a. the outside option, has a single terminal state with deterministic value* y. *A policy for the local game advances* $\mathcal{M}$ *for some (zero or non-zero) number of steps and either accepts the deterministic option* y *or the value from* $\mathcal{M}$. *Let* $f_{\mathcal{M}}(y)$ *denote the expected cost of the optimal policy for the local game* $(\mathcal{M}, y)$. *We refer to the function* $f_{\mathcal{M}}$ *as the optimality curve of the MDP* $\mathcal{M}$.

The surrogate cost of the outside option y (under any cost amortization) is simply y; as a consequence of Theorem 4.4, we obtain the following characterization:

COROLLARY 4.7. *For any Markov chain* $\mathcal{M}$ *and any* $y \in \mathbb{R}$, *it holds that* $f_{\mathcal{M}}(y) = \mathrm{E}\left[\min(y, W^*_{\mathcal{M}})\right]$.

Observe from this characterization that the CDF of the surrogate cost $W^*_{\mathcal{M}}$ can be derived[5] from the optimality curve as $1 - \frac{d}{dy} f_{\mathcal{M}}(y)$. Inspired by this connection, we can extend the definition of water filling surrogate costs to arbitrary MDPs:

DEFINITION 4.8. (WATER FILLING SURROGATE COSTS) *Let $\mathcal{M}$ be an MDP with optimality curve $f_{\mathcal{M}}$. The water filling surrogate cost for $\mathcal{M}$ is the random variable $W^*_{\mathcal{M}}$ generated from the CDF $1 - \frac{d}{dy} f_{\mathcal{M}}(y)$. That is, $W^*_{\mathcal{M}}$ is the random variable satisfying $f_{\mathcal{M}}(y) = \mathrm{E}\left[\min(y, W^*_{\mathcal{M}})\right]$ for all $y \in \mathbb{R}$.*

Our eventual goal is to prove the following analog of Theorem 4.4 for general CICS over MDPs.

THEOREM 4.9. *In any instance $\mathbb{I} = (\mathbb{M}, \mathcal{F})$ of matroid-min-CICS,* $\mathrm{OPT}(\mathbb{I}) \geq \mathrm{E}\left[\min_{S \in \mathcal{F}} \sum_{i \in S} W^*_{\mathcal{M}_i}\right]$.

The quantity $\mathrm{E}\left[\min_i W^*_{\mathcal{M}_i}\right]$ is called the Whittle integral in the context of single selection over infinite-horizon discounted-reward MDPs. It was first suggested as a lower bound on the optimal cost by [Whittle, 1980] and proved in that context by [Brown and Smith, 2013] and [Hadfield-Menell and Russell, 2015]. Unfortunately, Definition 4.8 does not give us insight into how the surrogate costs $W^*_{\mathcal{M}_i}$ relate to the cost of an adaptive algorithm for matroid-min-CICS. We will instead prove the theorem by relating optimality curves to the water filling procedure.

4.3 Water Filling and Surrogate Costs for General MDPs. In this section we will prove Theorem 4.9. In the case of Markov chains, the water filling surrogate cost of a randomly sampled trajectory recovers the optimality curve of the Markov chain. The challenge to extending this argument for general MDPs is that each sequence of actions creates a different distribution over the trajectories and consequently also over the surrogate costs, as illustrated in Example 4.5.

Our main observation is that if we are willing to cover action costs only partially through the amortization, then we can define an appropriate amortization for *every possible* deterministic commitment in $\mathcal{M}$ such that the distribution over surrogate costs generated by that commitment is independent of the commitment itself. In other words, *every* sequence of actions in the MDP generates the *same* distribution over surrogate costs. This distribution is a fundamental property of the MDP itself, and not of the algorithm operating on the MDP. We formalize this property through the following lemma.

LEMMA 4.10. *For any MDP $\mathcal{M} = (S, \sigma, A, c, \mathcal{D}, V, T)$ and any deterministic commitment $\pi \in \mathcal{C}(\mathcal{M})$, generating a Markov chain $\mathcal{M}^\pi$ with states $S_\pi \subseteq S$, realizable trajectories $\Upsilon_\pi \subseteq \Upsilon$ and a distribution p_π over trajectories and reachable states, there exists an amortized cost function $\rho^\pi : \Upsilon_\pi \mapsto \Delta(\mathbb{R})$ mapping trajectories to distributions over costs and a non-negative cost sharing vector $b^\pi = \{b^\pi_{s\tau}\}_{s \in S_\pi, \tau \in \Upsilon_\pi(s)}$, such that the following properties hold:*

1. **Cost Sharing:** *the amortized cost of a trajectory pays for its own acceptance value and the cost shares it sends to upstream states. For all $\tau \in \Upsilon_\pi$, it holds that $\mathrm{E}\left[\rho^\pi(\tau)\right] = v(\tau) + \sum_{s \in \tau} b^\pi_{s\tau_s}$.*

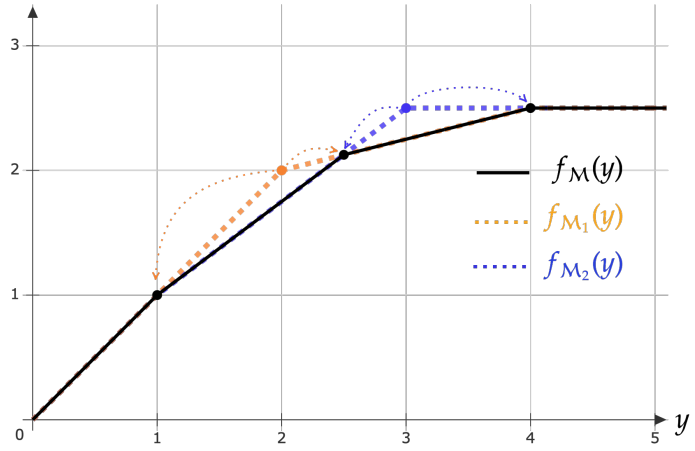
2. **Cost Dominance:** *the cost shares received by any state pay towards its action cost, but do not overpay. For all $s \in S_\pi$, it holds that $\sum_{\tau \in \Upsilon_\pi(s)} p_\pi(\tau) b^\pi_{s\tau} \leq p_\pi(s) c(\pi(s))$.*

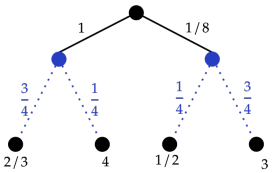
3. **Action Independence:** *sampling a trajectory $\tau \sim p_\pi$ and then sampling from the amortized cost distribution $\rho^\pi(\tau)$ generates a random surrogate cost for the MDP. This random variable is distributed identically to the water filling surrogate cost $W^*_{\mathcal{M}}$.*

As mentioned earlier, the amortized costs $\rho^\pi(\tau)$ defined by the lemma are necessarily different from the water filling costs in the respective Markov chains $\mathcal{M}^\pi$ as they must satisfy action independence. Observe that in the absence of the last condition connecting the surrogate costs $\rho^\pi(\tau)$ to the water filling costs $W^*_{\mathcal{M}}$ as defined in Definition 4.8, the lemma is trivial to satisfy. Indeed we can define all of the cost shares to be 0 and still satisfy cost dominance and action independence. The key contribution of the lemma is to show that we can achieve these properties while recovering the "optimal" surrogate costs as determined by the optimality curve $f_{\mathcal{M}}$. Figure 3 illustrates how the lemma applies to the MDP in Example 4.5.

[5] In particular, let H and h denote the CDF and PDF of $W^*_{\mathcal{M}}$ respectively, then we have $f_{\mathcal{M}}(y) = y(1 - H(y)) + \int_0^y z h(z) dz$, from which the statement follows by differentiating both sides.

Once we establish the lemma, the proof of Theorem 4.9 follows in much the same way as the proof of Lemma 4.2. At a high level, we can view the algorithm's choices *in hindsight* as corresponding to some deterministic commitment, and relate the algorithm's cost to the surrogate costs as instantiated through that commitment.





	t_{11}	t_{12}	t_{21}	t_{22}
$p(t)$	3/4	1/4	1/4	3/4
$v(t)$	2/3	4	1/2	3
$W^*_{\mathcal{M}_1}(t)$	2	4		
$W^*_{\mathcal{M}_2}(t)$			1	3
$\rho_{\mathcal{M}}(t)$	1 w.p. 1/3 5/2 w.p. 2/3	4	1	5/2 w.p. 2/3 4 w.p. 1/3

Figure 3: We consider the same setting as in Example 4.5. The optimality curve for MDP $\mathcal{M}$ is obtained by the minimum of the optimality curves for $\mathcal{M}_1$ (orange dashed line) and $\mathcal{M}_2$ (blue dashed line). Each action/commitment leads to two possible trajectories. The table on the right displays the four trajectories along with their respective probabilities of instantiation, values, and amortized costs under their respective commitments (commitment 1 for the first two columns and commitment 2 for the last two columns). Observe that resulting surrogate costs have different distributions. The last row in the table defines an alternate set of (random) surrogate costs $\rho_{\mathcal{M}}(t)$. Observe that for each t, we have $\mathrm{E}[\rho_{\mathcal{M}}(t)] \leq v(t)$. This implies cost dominance. The distribution of $\rho_{\mathcal{M}}(t)$ when t is picked according to commitment 1 is 1 with probability $(3/4)\cdot(1/3) = (1/4)$, 2.5 with probability $(3/4)\cdot(2/3) = (1/2)$ and 4 with probability $(1/4)\cdot(1) = (1/4)$. We obtain the same distribution if t is picked according to commitment 2. The same distribution also arises from the solid black optimality curve in the graph on the left, exhibiting the action independence property.

4.4 Omitted Proofs. We conclude this section by presenting the formal proofs of all our stated theorems and lemmas.

Proof of Lemma 4.2. We begin with the proof of the lower bound we stated in Lemma 4.2, relating the cost of any algorithm to the surrogate costs of any amortization. We note that one way to prove Lemma 4.2 is to follow the approach developed by [Singla, 2017] and relate the performance of any algorithm on a "costly information" instance to its performance in a "free information world" where action costs are paid by an outside investor who is in turn paid back the extra amortized cost of any terminal that the algorithm follows and accepts. Instead, we provide an algorithmic proof that will allow us to directly argue about the optimality of the water filling index policy, as well as extend our setting to MDPs.

Let ALG be any algorithm for $\mathbb{I}$. We will decompose the expected cost of ALG into the expected cost that it pays in every Markov chain $\mathcal{M}_i$ for all $i \in [n]$. Let $\mathcal{M}_i = (S_i, \sigma_i, A_i, c_i, \mathcal{D}_i, V_i, T_i)$ be the i-th Markov chain and let $\rho_i := \rho_{\mathcal{M}_i, b_i}$ be any cost amortization for this chain. We drop the i-index for notational convenience and define the following random events:

- $I(s) :=$ the event that ALG advances $\mathcal{M}$ at the non-terminal state $s \in S$.
- $R(\tau) :=$ the event that nature realizes a trajectory $\tau' \in \Upsilon$ with suffix τ.
- $R(s) :=$ the event that nature realizes a trajectory $\tau \in \Upsilon$ with $s \in \tau$.
- $A :=$ the event that ALG accepts a terminal state $t \in T$.

Then, the expected cost paid by the algorithm on Markov chain $\mathcal{M}$ can be written as:

$$\mathrm{ALG}(\mathcal{M}) = \sum_{s\in S} \Pr\left[I(s)\right] \cdot c(s) + \sum_{\tau\in\Upsilon} \Pr\left[R(\tau)\right] \cdot \Pr\left[A|R(\tau)\right] \cdot v(\tau) \tag{4.1}$$

For the expected cost of taking costly actions, we have that

$$\sum_{s\in S} \Pr\left[I(s)\right] \cdot c(s) = \sum_{s\in S} \Pr\left[I(s)\right] \cdot \sum_{\tau\in\Upsilon(s)} \frac{\Pr\left[R(\tau)\right]}{\Pr\left[R(s)\right]} \cdot b_{s\tau} = \sum_{s\in S} \Pr\left[I(s)\right] \cdot \sum_{\tau\in\Upsilon(s)} \Pr\left[R(\tau)|R(s)\right] \cdot b_{s\tau} \tag{4.2}$$

where the first equality follows by the definition of cost shares and the second equality follows by $\tau' \in \Upsilon(s)$.

For the expected cost of accepting a terminal state, we have that

$$\sum_{\tau\in\Upsilon} \Pr\left[R(\tau)\right] \cdot \Pr\left[A|R(\tau)\right] \cdot v(\tau) = \sum_{\tau\in\Upsilon} \Pr\left[R(\tau)\right] \cdot \Pr\left[A|R(\tau)\right] \cdot \left(\rho(\tau) - \sum_{s\in\tau} b_{s\tau_s}\right) \tag{4.3}$$

by definition of the surrogate cost $\rho(\tau)$. We further decompose this expression into two terms:

- For the first term, we have

$$\sum_{\tau\in\Upsilon} \Pr\left[R(\tau)\right] \cdot \Pr\left[A|R(\tau)\right] \cdot \rho(\tau) = \sum_{\tau\in\Upsilon} \Pr\left[R(\tau)\right] \cdot \mathrm{E}\left[A \cdot \rho(\tau)|R(\tau)\right] = \mathrm{E}\left[A \cdot \rho\right] \tag{4.4}$$

  with the first equality following from $A \in \{0,1\}$ and the second by definition of the surrogate cost.

- For the second term, we have

$$\begin{aligned} \sum_{\tau\in\Upsilon} \Pr\left[R(\tau)\right] \cdot \Pr\left[A|R(\tau)\right] \cdot \sum_{s\in\tau} b_{s\tau_s} &= \sum_{s\in S} \sum_{\tau\in\Upsilon(s)} b_{s\tau} \cdot \sum_{\tau'\in\Upsilon:\tau\subset\tau'} \Pr\left[R(\tau')\right] \cdot \Pr\left[A|R(\tau')\right] \\ &= \sum_{s\in S} \sum_{\tau\in\Upsilon(s)} b_{s\tau} \cdot \Pr\left[A \cap R(\tau)\right] \end{aligned} \tag{4.5}$$

  with the first line following by exchanging the summation order and the second line following by definition of the event $R(\tau)$, specifically that $R(\tau) = \sum_{\tau'\in\Upsilon:\tau\subset\tau'} R(\tau')$.

Combining Equations (4.1) through (4.5), we obtain that the expected cost of the algorithm on Markov chain $\mathcal{M}$ can be written as

$$\mathrm{ALG}(\mathcal{M}) = \mathrm{E}\left[A \cdot \rho\right] + \sum_{s\in S} \sum_{\tau\in\Upsilon(s)} b_{s\tau} \cdot \Big(\Pr\left[I(s)\right] \cdot \Pr\left[R(\tau)|R(s)\right] - \Pr\left[A \cap R(\tau)\right] \Big).$$

Observe that the second term is always non-negative: in order to accept a terminal from the chain **and** for nature to realize a trajectory $\tau \in \rho(s)$, it is necessary for the algorithm to advance state s (this already requires that s belongs in the realized trajectory) **and** for the suffix τ to be realized conditioned on having reached s; notice that the last two events are independent. Thus, we have concluded that

$$\mathrm{ALG}(\mathcal{M}) \geq \mathrm{E}\left[A \cdot \rho\right].$$

The proof is completed by noting that

$$\mathrm{E}\left[\mathrm{cost}(\mathrm{ALG})\right] = \sum_{i=1}^{n} \mathrm{ALG}(\mathcal{M}_i) \geq \sum_{i=1}^{n} \mathrm{E}\left[A_i \cdot \rho_i\right] = \mathrm{E}\left[\sum_{i=1}^{n} A_i \rho_i\right] \geq \mathrm{E}\left[\min_{S\in\mathcal{F}} \sum_{i\in S} \rho_i\right]$$

with the last inequality obtained by the fact that the set $S = \{i \in [n] : A_i = 1\}$ must be feasible (i.e. $S \in \mathcal{F}$) with probability 1 for ALG to be a valid policy. ∎

Proof of Theorem 4.4. We now proceed to prove Theorem 4.4 using Lemma 4.2. Observe that in the proof of Lemma 4.2, we only used inequalities at two points: (i) when relating the amortized cost of the trajectories that the algorithm selected to the (feasible) set of trajectories of minimum total surrogate cost and (ii) when we established that $\Pr[I(s)] \cdot \Pr[R(\tau)|R(s)] \geq \Pr[A \cap R(\tau)]$ for all states s and all trajectories $\tau \in \Upsilon(s)$; notice that this was only required for pairs (s, τ) with $b_{s\tau} > 0$. The first inequality can become tight if the algorithm somehow ensures that it will always accept the feasible set of realized trajectories whose total surrogate cost is minimum. The second inequality becomes tight if the algorithm can somehow ensure that whenever it advances a state s sharing cost $b_{s\tau} > 0$ with one of its downstream trajectories τ, it will always accept the corresponding Markov chain if the suffix τ gets realized. This allows us to characterize the conditions under which the lower bound is actually met by an algorithm, which we formally state as the following corollary:

COROLLARY 4.11. *Consider any instance $\mathbb{I} = (\mathbb{M}, \mathcal{F})$ of matroid-min-CICS over Markov chains and let b_i be any cost amortization of $\mathcal{M}_i$ with surrogate cost $\rho_i := \rho_{\mathcal{M}_i, b_i}$ for all $i \in [n]$. Then, any algorithm that satisfies both*

1. ***Surrogate Optimality.*** *The algorithm always accepts a feasible set of Markov chains whose realized trajectories have the minimum total surrogate cost.*
2. ***Promise of Payment.*** *Whenever one of the Markov chains $\mathcal{M}_i$ gets advanced from a state $s_i \in S_i$ such that $b_{s_i\tau_i} > 0$ for some $\tau_i \in \Upsilon(s_i)$, the algorithm will keep advancing $\mathcal{M}_i$ as long as it's trajectory evolves according to τ_i.*

will have expected cost precisely $\mathrm{E}\left[\min_{S \in \mathcal{F}} \sum_{i \in S} \rho_i\right]$.

Promise of payment states that whenever $b_{s\tau} > 0$ and the algorithm advances s, it will ensure that while it is possible for τ to be realized, the corresponding Markov chain will continue to be advanced; in other words, that $\Pr[I(s)] \cdot \Pr[R(\tau)|R(s)] = \Pr[A \cap R(\tau)]$. We note that achieving any of these properties individually is trivial; fully advancing all Markov chains and picking the feasible set of trajectories of minimum total surrogate cost achieves surrogate optimality, and either advancing a Markov chain fully or not advancing it at all achieves promise of payment. The water filling amortization is specifically defined so that the corresponding index policy achieves both of these properties, simultaneously.

It suffices to argue that the water filling index policy, paired with the water filling amortization, satisfies both surrogate optimality and promise of payment. We begin with surrogate optimality. By definition, the water filling index policy advances the chain of minimum water filling index; recall that the index of a state corresponds to the minimum surrogate cost among all trajectories passing through it. This ensures that while there is potential for some chain $\mathcal{M}_i$ to realize the trajectory of minimum surrogate cost, the algorithm will keep advancing it. In other words, this algorithms ends up **greedily** accepting Markov chains with respect to the surrogate cost of their realized trajectories. Paired with the fact that a minimum cost basis of a matroid is always obtained by greedily adding the cheaper feasible element, this establishes surrogate optimality of the water filling index policy.

Next, we show that promise of payment also holds. Say that at any point the algorithm advances Markov chain $\mathcal{M}_i$ at state s_i with index $I_i(s_i)$ and there exists some trajectory $\tau_i \in \Upsilon(s_i)$ such that $b_{s_i\tau_i} > 0$. We need to argue that as long as the state of $\mathcal{M}_i$ evolves according to τ_i, the algorithm will keep advancing it. Since $\mathcal{M}_i$ was advanced by the water filling index policy, this implies that $I_i(s_i)$ was the minimum index among the current states of all Markov chains; thus, a sufficient condition is to show that $I_i(s_i') \leq I_i(s_i)$ for all $s_i' \in \tau_i$, as then all the states in the trajectory will continue to have the minimum index and thus will be advanced if realized. Since $b_{s_i\tau_i} > 0$, we know by definition of the water filling amortization that immediately after s_i got amortized, τ_i had the minimum surrogate cost across all trajectories in $\Upsilon(s_i)$. As states are considered in a bottom-up order, this implies that any subsequent steps of the amortization will maintain that the trajectory passing through s_i that has the minimum surrogate cost ends with τ_i. This immediately gives us the proof, by definition of the indices and $s_i' \in \tau_i$. ∎

Proof of Lemma 4.10. We now proceed to prove Lemma 4.10, which is the main technical challenge of this section. Fix any MDP $\mathcal{M} = (S, \sigma, A, c, \mathcal{D}, V, T)$. The proof follows by induction on the horizon H of the MDP. If $H = 1$, then $\mathcal{M}$ is a singleton MDP where $\sigma \in T$ is the unique state, π_σ is the unique (trivial) commitment and $\Upsilon_{\pi_\sigma} = \{\{\sigma\}\}$. Therefore, the lemma is immediately satisfied by $\rho^{\pi_\sigma}(\{\sigma\}) := v(\sigma)$ and $b^{\pi_\sigma}_{\sigma\{\sigma\}} := 0$. We now assume that the conditions of the lemma hold for all MDPs of horizon up to H and extend it to MDPs of horizon $H + 1$.

Let $\mathcal{M}$ be of horizon $H+1$ and let $A(\sigma) = \{a_1, \cdots, a_k\}$ be the set of actions available at the starting state σ. Let R_j denote the set of all states that can be reached directly by taking action a_j on σ; that is, $R_j := \{s \in S : \mathcal{D}(\sigma, a_j, s) > 0\}$. For each $s \in R_j$, we use $\mathcal{M}_s$ to denote the subprocesses of $\mathcal{M}$ that starts in state s and $W^*_{\mathcal{M}_s}$ to denote the water filling surrogate cost of $\mathcal{M}_s$. We begin by computing the water filling amortization of each action a_j.

Let Z_j denote the random variable that first draws a state s from the distribution $\mathcal{D}(\sigma, a_j, \cdot)$ and then draws a value from the distribution $W^*_{\mathcal{M}_s}$. Let g_j be the solution to the equation

$$c(a_j) + \mathrm{E}\left[Z_j\right] = \mathrm{E}\left[\max\{g_j, Z_j\}\right].$$

Then, $\hat{Z}_j := \max\{g_j, Z_j\}$ is the water filling surrogate cost of the action a_j. We observe that for any $y \geq g_j$, we have $c(a_j) + \mathrm{E}\left[\min\{y, Z_j\}\right] = \mathrm{E}\left[\min\{y, \hat{Z}_j\}\right]$, and because $\hat{Z}_j$ is always at least g_j, we get for all y:

$$\min\{y, c(a_j) + \mathrm{E}\left[\min\{y, Z_j\}\right]\} = \mathrm{E}\left[\min\{y, \hat{Z}_j\}\right]. \tag{4.6}$$

We can now write the optimality curve of $\mathcal{M}$ as:

$$\begin{aligned} f_{\mathcal{M}}(y) &= \min\left\{y, \min_{j\in[k]}\left(c(a_j) + \sum_{s\in R_j} \mathcal{D}(\sigma, a_j, s)\cdot f_{\mathcal{M}_s}(y)\right)\right\} \\ &= \min\left\{y, \min_{j\in[k]}\left(c(a_j) + \sum_{s\in R_j} \mathcal{D}(\sigma, a_j, s)\cdot \mathrm{E}\left[\min\{y, W^*_{\mathcal{M}_s}\}\right]\right)\right\} \\ &= \min\left\{y, \min_{j\in[k]}\left(c(a_j) + \mathrm{E}\left[\min\{y, Z_j\}\right]\right)\right\} = \min\left\{y, \min_{j\in[k]} \mathrm{E}\left[\min\left(y, \hat{Z}_j\right)\right]\right\}. \end{aligned}$$

Here the first equation follows from noting that in the local game $(\mathcal{M}, y)$, the algorithm can either choose the outside option y or takes one of the actions a_j from σ and then proceeds optimally in the game $(\mathcal{M}_s, y)$ where s is instantiated from R_j. The second equation follows by the definition of $W^*_{\mathcal{M}_s}$; the third by the definition of Z_j; and the fourth by Equation (4.6).

Recall that $f_{\mathcal{M}}(y) = \mathrm{E}\left[\min\{y, W^*_{\mathcal{M}}\}\right]$, and so, we conclude that

$$\mathrm{E}\left[\min\{y, W^*_{\mathcal{M}}\}\right] \leq \mathrm{E}\left[\min\{y, \hat{Z}_j\}\right]$$

for all $j \in [k]$. This implies that the random variable $\hat{Z}_j$ second-order stochastically dominates the random variable $W^*_{\mathcal{M}}$, allowing us to use the following lemma.

Lemma 4.12. (Second Order Stochastic Dominance.) *Let X, Z be discrete random variables that satisfy the property $\mathrm{E}\left[\min\{y, X\}\right] \leq \mathrm{E}\left[\min\{y, Z\}\right]$ for all $y \in \mathbb{R}$. There exists a mapping $m : \mathrm{supp}(Z) \mapsto \Delta(\mathrm{supp}(X))$ from the support of Z to distributions over the support of X such that:*

1. *X is obtained by sampling from $m(z)$ for a randomly sampled $z \sim Z$.*
2. *For all $z \in \mathrm{support}(Z)$, it holds that $\mathrm{E}\left[m(z)\right] \leq z$.*

We note that the lemma is standard (see for example [Strassen, 1965, Föllmer and Schied, 2016]) but we provide a constructive proof in Appendix D for the sake of intuition and completeness. We apply Lemma 4.12 to all tuples $(W^*_{\mathcal{M}}, \hat{Z}_j)$ to obtain mappings $m_j(\cdot)$.

We are finally ready to define the amortized cost functions ρ^π and the cost sharing vectors b^π. Fix any deterministic commitment $\pi \in \mathcal{C}\left(\mathcal{M}\right)$ and let $j = \pi(\sigma) \in [k]$ be the fixed action that π takes at state σ. For each state $s \in R_j$, we use $\pi|s \in \mathcal{C}\left(\mathcal{M}_s\right)$ to denote the commitment π after we transition to state s; note that this is also a deterministic commitment. By definition, each MDP $\mathcal{M}_s$ has horizon up to H and thus by the induction hypothesis, it admits a mapping $\rho^{\pi|\sigma}(\cdot)$ and a non-negative cost sharing vector $b^{\pi|\sigma}$, satisfying the properties

of the lemma. Now, consider any trajectory $\tau \in \Upsilon_\pi$ and observe that $\tau = \{\sigma, \tau_s\}$ for some $s \in R_j$ and some $\tau_s \in \Upsilon_\pi(s)$. We define

$$\rho^\pi(\tau) := m_j(\max\{g_j, \rho^{\pi|s}(\tau_s)\})$$

and

$$b^\pi_{\sigma\tau} := \mathrm{E}\left[\rho^\pi(\tau)\right] - \mathrm{E}\left[\rho^{\pi|s}(\tau_s)\right]$$

and for all other $s \in S_\pi \setminus \{\sigma\}$ and $\tau \in \Upsilon_\pi(s)$, we use the same cost share $b^\pi_{s\tau} = b^{\pi|s}_{s\tau}$ that was used in $\mathcal{M}_s$.

- **Cost sharing:** This holds trivially by the definition of the cost shares in the starting state σ, the fact that we don't change the cost shares in any other state $s \neq \sigma$, the fact that $v(\tau) = v(\tau_s)$ for any suffix τ_s of τ and the induction hypothesis.

- **Cost dominance:** By the induction hypothesis and the fact that we maintain the cost shares of every state $s \neq \sigma$, we only need to prove the inequality for σ. We have

$$\begin{aligned}
\sum_{\tau\in\Upsilon_\pi} p_\pi(\tau) b^\pi_{\sigma\tau} &= \sum_{s\in R_j, \tau_s \in \Upsilon_\pi(s)} p_\pi(\{\sigma,\tau_s\})(\mathrm{E}\left[\rho^\pi(\{\sigma,\tau_s\})\right] - \mathrm{E}\left[\rho^{\pi|s}(\tau_s)\right]) \\
&= \sum_{\tau\in\Upsilon_\pi} p^\pi(\tau)\,\mathrm{E}\left[\rho^\pi(\tau)\right] - \sum_{s\in R_j} \mathcal{D}(\sigma, a_j, s) \sum_{\tau_s\in\Upsilon_\pi(s)} p^{\pi|s}(\tau_s)\,\mathrm{E}\left[\rho^{\pi|s}(\tau_s)\right] \\
&= \mathrm{E}\left[W^*_{\mathcal{M}}\right] - \sum_{s\in R_j} \mathcal{D}(\sigma, a_j, s)\,\mathrm{E}\left[W^*_{\mathcal{M}_s}\right] \\
&\leq \mathrm{E}\left[\hat{Z}_j\right] - \mathrm{E}\left[Z_j\right] = c(a_j).
\end{aligned}$$

  and the statement follows by noting $p^\pi(\sigma) = 1$. Here the first equation follows from the definition of $b^\pi_{\sigma\tau}$ and a decomposition of the trajectories τ; the second just rewrites the terms separately; the third is by the action independence of ρ^π (proven below), and by the induction hypothesis applied to $\mathcal{M}_s$ similarly; the fourth uses property (2) in Lemma 4.12 for the first term and the definition of Z_j for the second term; and the last equality follows from the definition of $\hat{Z}_j$.

- **Action independence:** Drawing a trajectory $\tau \sim p_\pi$ is equivalent to first drawing a state $s \in R_j$ from the distribution $\mathcal{D}(\sigma, a_j, \cdot)$ and then drawing a trajectory $\tau_s \in \Upsilon_\pi(s)$ from $p_{\pi|s}$. Consider drawing τ in this manner and then sampling from the distribution $\rho^{\pi|s}(\tau_s)$. By the induction hypothesis and the definition of Z_j, this provides us with a sample drawn from Z_j. Then, $\max\{g_j, \rho^{\pi|s}(\tau_s)\}$ with τ drawn in this manner corresponds to an instantiation of $\hat{Z}_j$, and m_j applied to that instantiation results in an instantiation of $W^*_{\mathcal{M}}$ by the definition of m_j and property (1) in Lemma 4.12.

This concludes the proof of Lemma 4.10. ∎

Proof of Theorem 4.9. Finally, we are ready to prove our main result (Theorem 4.9) using Lemma 4.10. Fix any algorithm (including the optimal adaptive policy) for instance $\mathbb{I}$ and let ALG denote the (random) cost of this algorithm. Also, for all $i \in [n]$, let ALG(i) denote the (random) cost suffered by the algorithm from paying action costs and accepting terminal states in $\mathcal{M}_i$. Notice that ALG $= \sum_{i=1}^n$ ALG(i) with probability 1. Finally, let $X(i)$ be the random variable indicating whether the algorithm accepts a state from $\mathcal{M}_i$ or not. To prove Theorem 4.9, we will argue that for all $i \in [n]$,

$$\mathrm{E}\left[\mathrm{ALG}(i)\right] \geq \mathrm{E}\left[X(i) \cdot W^*_{\mathcal{M}_i}\right]. \tag{4.7}$$

Notice that if this is true, then we immediately have that

$$\mathrm{E}\left[\mathrm{ALG}\right] = \sum_{i=1}^n \mathrm{E}\left[\mathrm{ALG}(i)\right] \geq \sum_{i=1}^n \mathrm{E}\left[X(i) \cdot W^*_{\mathcal{M}_i}\right] = \mathrm{E}\left[\sum_{i=1}^n X(i) \cdot W^*_{\mathcal{M}_i}\right] \geq \mathrm{E}\left[\min_{S\in\mathcal{F}} \sum_{i\in S} W^*_{\mathcal{M}_i}\right]$$

with the last inequality following from the fact that for any feasible algorithm, the set of accepted terminals $S = \{i : X(i) = 1\}$ must be feasible (i.e. $S \in \mathcal{F}$) with probability 1.

We now turn our attention to proving inequality (4.7). We use $\mathcal{A}$ to denote the inner randomness of the algorithm and $\mathcal{R}_i$ to denote the randomness of each MDP $\mathcal{M}_i$ for all $i \in [n]$. Notice that conditioned on $\mathcal{A}$ and $\mathcal{R}_i$ for all $i \in [n]$, the outcome of the algorithm is deterministic. Now fix any $i \in [n]$ and let

$$\mathcal{R}^{-i} := \mathcal{A} \cup \left(\cup_{j \neq i} \mathcal{R}_j \right)$$

encode all the randomness in the algorithm's run **except** from the realizations of $\mathcal{M}_i$. The key observation is that conditioned on $\mathcal{R}^{-i}$, all the actions that the algorithm takes in MDP $\mathcal{M}_i$ are *predetermined*; in other words, the algorithm's trajectory on $\mathcal{M}_i$ is fully described by some deterministic commitment denoted $\pi_i = \pi_i(\mathcal{R}^{-i}) \in \mathcal{C}(\mathcal{M}_i)$. Thus, we have that

$$\mathrm{E}[\mathrm{ALG}(i)] = \mathrm{E}_{\mathcal{R}^{-i}}\left[\mathrm{E}_{\mathcal{R}_i}\left[\mathrm{ALG}(i)|R^{-i}\right]\right] = \mathrm{E}_{\mathcal{R}^{-i}}\left[\mathrm{cost}(\pi_i)\right] \tag{4.8}$$

where $cost(\pi_i)$ is the expected cost of the algorithm, following the commitment $\pi_i = \pi_i(\mathcal{R}^{-i})$, on MDP $\mathcal{M}_i$. The proof is the completed by the following generalization of Lemma 4.2, which is enabled from Lemma 4.10.

Claim 4.13. *Fix any MDP $\mathcal{M}$ and any deterministic commitment $\pi \in \mathcal{C}(\mathcal{M})$. Then, the expected cost of any algorithm following π on $\mathcal{M}$ will be at least*

$$\mathrm{E}[X(\pi) \cdot W^*_{\mathcal{M}}]$$

where $X(\pi)$ is an indicator of whether the algorithm accepts a terminal state of $\mathcal{M}$ or not.

Notice that the above claim doesn't depend on the underlying CICS instance $\mathbb{I}$; it simply states that conditioned on running a committing policy on some MDP and accepting a terminal state, the expected total cost spent on this MDP is lower bounded by the surrogate cost. Since this lower bound applies to all commitments $\pi \in \mathcal{C}(\mathcal{M})$, coupled with equation (4.8), it directly implies inequality (4.7), as

$$\mathrm{E}[\mathrm{ALG}(i)] = \mathrm{E}_{\mathcal{R}^{-i}}[\mathrm{cost}(\pi_i)] \geq \mathrm{E}_{R^{-i}}\left[\mathrm{E}\left[(X(i)|R^{-i}) \cdot W^*_{\mathcal{M}_i}\right]\right] = \mathrm{E}\left[X(i) \cdot W^*_{\mathcal{M}_i}\right],$$

completing the proof of Theorem 4.9. ∎

Proof of Claim 4.13. Notice that since we are committing to running policy $\pi \in \mathcal{C}(\mathcal{M})$ on $\mathcal{M}$, we are essentially running some algorithm on the Markov chain $\mathcal{M}^\pi$. From Lemma 4.2 and Theorem 4.4, we know that the contribution of Markov chain $\mathcal{M}^\pi$ to the total cost will be at least

$$\mathrm{E}[X(\pi) \cdot W^*_{\mathcal{M}^\pi}]$$

and thus to prove the claim, we will need to show that

$$\mathrm{E}[W^*_{\mathcal{M}^\pi}] \geq \mathrm{E}[W^*_{\mathcal{M}}]$$

for all $\pi \in \mathcal{C}(\mathcal{M})$.

Let $\mathcal{M} = (S, \sigma, A, c, \mathcal{D}, V, T)$ and recall that we use S_π and Υ_π to denote the state space and trajectory set of $\mathcal{M}^\pi$ and p_π to denote the implied distribution over Υ_π. Finally, we use $c(s)$ to denote the cost of the unique action that the deterministic commitment π chooses at state $s \in S_\pi$. By definition of the water filling surrogate cost of a Markov chain, we have

$$\mathrm{E}[W^*_{\mathcal{M}^\pi}] = \sum_{\tau \in \Upsilon_\pi} p_\pi(\tau) \cdot \rho^*_{\mathcal{M}^\pi}(\tau) = \sum_{\tau \in \Upsilon_\pi} p_\pi(\tau) \cdot \left(v(\tau) + \sum_{s \in \tau} b^*_{s\tau_s} \right)$$

where $b^*_{s\tau}$ are non-negative cost shares satisfying $\sum_{\tau \in \Upsilon_\pi(s)} p_\pi(\tau) b^*_{s\tau} = p_\pi(s)c(s)$ for all $s \in S_\pi$. Since π is deterministic, we know from Lemma 4.10 that there exists an amortized cost function $\rho^\pi(\cdot)$ over the trajectories in Υ_π and a set of cost shares $\{b^\pi_{s\tau}\}$ that will satisfy action independence, cost sharing and cost dominance. Using these, we have

$$\mathrm{E}[W^*_{\mathcal{M}^\pi}] = \sum_{\tau \in \Upsilon_\pi} p_\pi(\tau) \cdot \left(v(\tau) + \sum_{s \in \tau} b^*_{s\tau_s} \right)$$

$$
\begin{aligned}
&= \sum_{\tau\in\Upsilon_\pi} p_\pi(\tau)\cdot\left(\mathrm{E}\left[\rho^\pi(\tau)\right]-\sum_{s\in\tau} b^\pi_{s\tau_s}+\sum_{s\in\tau} b^*_{s\tau_s}\right) && \text{(Cost Sharing)}\\
&= \sum_{\tau\in\Upsilon_\pi} p_\pi(\tau)\cdot\mathrm{E}\left[\rho^\pi(\tau)\right]+\sum_{\tau\in\Upsilon_\pi}\sum_{s\in\tau} p_\pi(\tau)(b^*_{s\tau_s}-b^\pi_{s\tau_s})\\
&= \mathrm{E}\left[W^*_{\mathcal{M}}\right]+\sum_{s\in S_\pi}\sum_{\tau\in\Upsilon_\pi(s)} p_\pi(\tau)(b^*_{s\tau}-b^\pi_{s\tau}) && \text{(Action Independence)}\\
&= \mathrm{E}\left[W^*_{\mathcal{M}}\right]+\sum_{s\in S_\pi} p_\pi(s)c(s)-\sum_{\tau\in\Upsilon_\pi(s)} p_\pi(\tau)b^\pi_{s\tau}\\
&\geq \mathrm{E}\left[W^*_{\mathcal{M}}\right]. && \text{(Cost Dominance)}
\end{aligned}
$$

□

5 Local Approximation and Composition Theorems

Recall that the commitment gap of a CICS instance is defined as the ratio between the cost of the optimal committing policy and the cost of the optimal (non-committing) policy. Let $\mathbb{I}=(\mathbb{M},\mathcal{F})$ be an instance of matroid-min-CICS and let π_i be a commitment for $\mathcal{M}_i\in\mathbb{M}$. We will establish bounds on the commitment gap achieved by $\mathcal{P}=(\pi_1,\cdots,\pi_n)$ by quantifying for each $i\in[n]$ the performance of π_i in the local game $(\mathcal{M}_i,y)$.

Recall that the performance of an optimal algorithm for the local game $(\mathcal{M},y)$ is described by the optimality curve $f_{\mathcal{M}}(y)$. The performance of the commitment π in the same game is given by $f_{\mathcal{M}^\pi}(y)$. [Scully and Doval, 2024] define local approximation by relating these two quantities:[6]

Definition 5.1. (Local Approximation) *Let $\mathcal{M}$ be any MDP and let $\alpha\geq 1$. We say that a commitment $\pi\in\mathcal{C}(\mathcal{M})$ is an α-local approximation for $\mathcal{M}$ if $f_{\mathcal{M}^\pi}(\alpha y)\leq\alpha\cdot f_{\mathcal{M}}(y)$ for all $y\in\mathbb{R}$.*

Restating the inequality in terms of the water-filling costs provides a more intuitive definition as well as a composition theorem. Theorems 4.4 and 4.9 together imply that we can bound the commitment gap as:

$$
\mathrm{CG}(\mathbb{I})\leq\min_{(\pi_1,\cdots,\pi_n)\in\mathcal{C}(\mathbb{I})}\frac{\mathrm{E}\left[\min_{S\in\mathcal{F}}\sum_{i\in S}W^*_{\mathcal{M}_i^{\pi_i}}\right]}{\mathrm{E}\left[\min_{S\in\mathcal{F}}\sum_{i\in S}W^*_{\mathcal{M}_i}\right]}.
$$

On the other hand, using Definition 4.8, an α-local approximation for π_i with respect to $\mathcal{M}_i$ can be restated as a second-order stochastic dominance condition:

$$
\mathrm{E}\left[\min\{y,W^*_{\mathcal{M}_i^{\pi_i}}\}\right]\leq\mathrm{E}\left[\min\{y,\alpha W^*_{\mathcal{M}_i}\}\right]\quad\forall y\in\mathbb{R}. \tag{5.9}
$$

Combining these inequalities provides the following composition result.

Theorem 5.2. *Let $\mathbb{I}=(\mathbb{M},\mathcal{F})$ be any instance of matroid-min-CICS, where each constituent MDP $\mathcal{M}_i$ admits an α-local approximation under some commitment $\pi_i\in\mathcal{C}(\mathcal{M}_i)$. Then, $\mathrm{CG}(\mathbb{I})\leq\alpha$.*

Proof. Let $z:=[W^*_{\mathcal{M}_2^\pi},\cdots W^*_{\mathcal{M}_n^\pi}]$ encode the surrogate costs of all Markov chains $\mathcal{M}_i^{\pi_i}$ with the exception of $\mathcal{M}_1^{\pi_1}$. Then, we have

$$
\begin{aligned}
\mathrm{E}\left[\min_{S\in\mathcal{F}}\sum_{i\in S}W^*_{\mathcal{M}_i^{\pi_i}}\right] &= \mathrm{E}_z\left[\mathrm{E}_{W^*_{\mathcal{M}_1^{\pi_1}}}\left[\min_{S\in\mathcal{F}}\sum_{i\in S}W^*_{\mathcal{M}_i^{\pi_i}}\right]\right]\\
&= \mathrm{E}_z\left[\mathrm{E}_{W^*_{\mathcal{M}_1^{\pi_1}}}\left[\min_{S\in\mathcal{F}}\left(W^*_{\mathcal{M}_1^{\pi_1}}\cdot\mathbb{1}[1\in S]+\sum_{i\in S\setminus\{1\}}W^*_{\mathcal{M}_i^{\pi_i}}\right)\right]\right]\\
&= \mathrm{E}_z\left[\mathrm{E}_{W^*_{\mathcal{M}_1^{\pi_1}}}\left[\min\left(\min_{S\in\mathcal{F}:1\notin S}\sum_{i\in S}W^*_{\mathcal{M}_i^{\pi_i}}\ ,\ W^*_{\mathcal{M}_1^{\pi_1}}+\min_{S:1\notin S;S\cup\{1\}\in\mathcal{F}}\sum_{i\in S}W^*_{\mathcal{M}_i^{\pi_i}}\right)\right]\right]
\end{aligned}
$$

[6] Their definition is provided in the context of optional-inspection PB, and we extend the same definition to general MDPs.

Now, let $f_1(z) := \min_{S\in\mathcal{F}:1\notin S}\sum_{i\in S} W^*_{\mathcal{M}_i^{\pi_i}}$ and $f_2(z) := \min_{S:1\notin S;S\cup\{1\}\in\mathcal{F}}\sum_{i\in S} W^*_{\mathcal{M}_i^{\pi_i}}$. Note that both quantities depend only on z and not on $W^*_{\mathcal{M}_1^{\pi_1}}$. Then, we have

$$\begin{aligned}
\mathrm{E}\left[\min_{S\in\mathcal{F}}\sum_{i\in S} W^*_{\mathcal{M}_i^{\pi_i}}\right] &= \mathrm{E}_z\left[\mathrm{E}_{W^*_{\mathcal{M}_1^{\pi_1}}}\left[\min\left(f_1(z)\ ,\ W^*_{\mathcal{M}_1^{\pi_1}}+f_2(z)\right)\right]\right] \\
&= \mathrm{E}_z\left[f_2(z)+\mathrm{E}_{W^*_{\mathcal{M}_1^{\pi_1}}}\left[\min\left(f_1(z)-f_2(z)\ ,\ W^*_{\mathcal{M}_1^{\pi_1}}\right)\right]\right] \\
&\le \mathrm{E}_z\left[f_2(z)+\mathrm{E}_{W^*_{\mathcal{M}_1}}\left[\min\left(f_1(z)-f_2(z)\ ,\ \alpha\cdot W^*_{\mathcal{M}_1}\right)\right]\right] \qquad \text{(Local Approximation)} \\
&= \mathrm{E}_z\left[\mathrm{E}_{W^*_{\mathcal{M}_1}}\left[\min\left(f_1(z)\ ,\ \alpha\cdot W^*_{\mathcal{M}_1}+f_2(z)\right)\right]\right] \\
&= \mathrm{E}\left[\min_{S\in\mathcal{F}}\sum_{i\in S}\tilde{W}_i\right]
\end{aligned}$$

where $\tilde{W}_1 = \alpha\cdot W^*_{\mathcal{M}_1}$ and $\tilde{W}_i := W^*_{\mathcal{M}_i^{\pi_i}}$ for all $i\neq 1$. Thus, we have substituted $W^*_{\mathcal{M}_1^{\pi_1}}$ with $\alpha\cdot W^*_{\mathcal{M}_1}$. By repeating the same process for $i = 2, 3, \cdots, n$, we obtain that

$$\mathrm{E}\left[\min_{S\in\mathcal{F}}\sum_{i\in S} W^*_{\mathcal{M}_i^{\pi_i}}\right] \le \mathrm{E}\left[\min_{S\in\mathcal{F}}\sum_{i\in S}\alpha\cdot W^*_{\mathcal{M}_i}\right]$$

and thus

$$\frac{\mathrm{E}\left[\min_{S\in\mathcal{F}}\sum_{i\in S} W^*_{\mathcal{M}_i^{\pi_i}}\right]}{\mathrm{E}\left[\min_{S\in\mathcal{F}}\sum_{i\in S} W^*_{\mathcal{M}_i}\right]} \le \alpha.$$

From Theorem 4.4 the nominator is precisely the cost of the optimal committing policy for $\mathbb{I}$ under commitments $\mathcal{P} = (\pi_1, \cdots, \pi_n)$ and from Theorem 4.9, the denominator is a lower bound on $\mathrm{OPT}(\mathbb{I})$. Thus, we have $\mathrm{CG}(\mathbb{I}) \le \alpha$ as desired. □

Theorem 5.2 also provides a recipe for designing approximation algorithms for CICS: if we can efficiently identify a commitment π_i for each MDP $\mathcal{M}_i$ that achieves an α-local approximation, then we can simply restrict each MDP $\mathcal{M}_i$ to the corresponding Markov chain $\mathcal{M}_i^{\pi_i}$ and run the (efficient) water filling index policy; the resulting committing policy is guaranteed to be an α-approximation to the unrestricted optimum.

Geometric Intuition for Local Approximation. It is worth emphasizing that α-local approximation is *not equivalent* to simply achieving a standard α-approximation in the local game. Slightly rearranging the condition in Definition 5.1, we have that for commitment π to be an α-local approximation, we require for all $y\in\mathbb{R}$,

$$f_{\mathcal{M}^\pi}(y) \le \alpha f_{\mathcal{M}}(y/\alpha).$$

In contrast, for commitment π to be a standard α-approximation for the local game, we require only

$$f_{\mathcal{M}^\pi}(y) \le \alpha f_{\mathcal{M}}(y),$$

the difference being y instead of y/α on the right-hand side. As illustrated in Figure 4, we can understand both of the above conditions as saying that the graph of $f_{\mathcal{M}^\pi}(y)$ needs to lie below a geometrically scaled version of the graph of $f_{\mathcal{M}}(y)$. However, the scaling used is different for the two conditions.

- For standard α-approximation, we vertically scale up $f_{\mathcal{M}}(y)$ by a factor of α.
- For α-local approximation, we vertically *and horizontally* scale up $f_{\mathcal{M}}(y)$ by a factor of α.

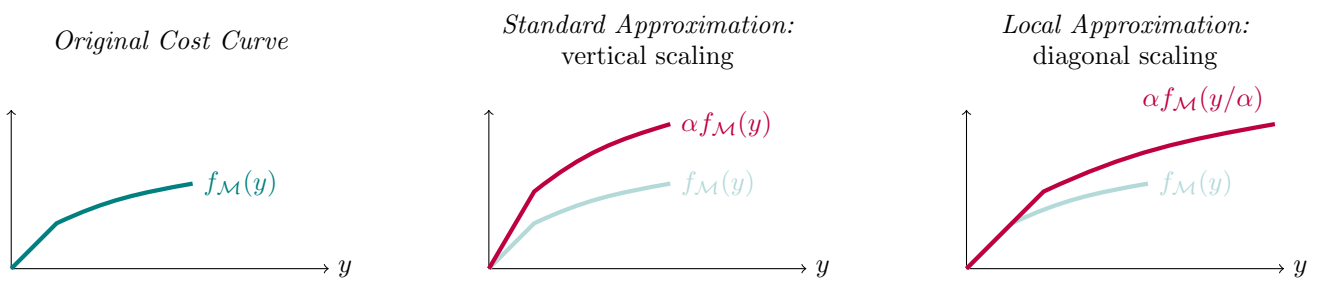


Figure 4: For an MDP $\mathcal{M}$ with a given optimality curve (left), we contrast two approximation notions: standard α-approximation (center) vs. α-local approximation (right). In both cases, achieving the approximation requires designing a commitment whose optimality curve lies below a scaled version of the optimal cost (purple curves). For standard α-approximation, the curve is scaled up vertically by a factor of α. For α-local approximation, the curve is scaled up *diagonally*, i.e. both vertically and horizontally, by a factor of α.

The additional horizontal scaling in local approximation is critical in proving eq. (5.9) and thereby obtaining Theorem 5.2. Standard α-approximation implies only a weaker version of eq. (5.9) with the y on the right-hand side replaced by αy. Working through the proof of Theorem 5.2 with this change, one would eventually obtain a commitment gap bound of α^n, scaling exponentially[7] in the number of MDPs n. Using the right local approximation notion is thus critical for obtaining approximation guarantees that do not depend on n.

Establishing Local Approximation. We note that in some cases, the surrogate costs of an MDP might not admit a simple structure, making second-order stochastic dominance conditions difficult to establish. We can alternatively establish local approximation through a stronger but conceptually easier relationship between the surrogate costs: namely, that $\alpha W^*_{\mathcal{M}}$ first-order stochastically dominates $W^*_{\mathcal{M}_\pi}$. Formally, we define the following alternate notion of approximation, where for a quantile $q \in [0,1]$ and a random variable X, we let $X(q)$ denote the qth quantile value of the r.v.: $X(q) = \inf\{x : Pr[X \le x] \ge q\}$.

Definition 5.3. (Pointwise Approximation) *Let $\mathcal{M}$ be any MDP and let $\alpha \ge 1$. We say that a commitment $\pi \in \mathcal{C}(\mathcal{M})$ is an α-pointwise approximation for $\mathcal{M}$ if $W^*_{\mathcal{M}^\pi}(q) \le \alpha \cdot W^*_{\mathcal{M}}(q)$ for all $q \in [0,1]$.*

The following is immediate from the fact that first-order stochastic dominance implies second-order stochastic dominance, but the converse need not hold.

Fact 5.4. *If $\pi \in \mathcal{C}(\mathcal{M})$ is an α-pointwise approximation for $\mathcal{M}$, it also an α-local approximation for $\mathcal{M}$.*

Finally, we note that by definition we have $f_{\mathcal{M}}(y) = \min_{\pi \in \mathcal{C}(\mathcal{M})} f_{\mathcal{M}^\pi}(y)$. Moreover, because randomized commitments generate optimality curves as averages over those of deterministic commitments, we can write $f_{\mathcal{M}}(y)$ as the minimum over the optimality curves of deterministic commitments. Thus, we can obtain local or pointwise approximation guarantees by showing that the commitment π satisfies the desired definition against *all other deterministic commitments* $\pi' \in \mathcal{C}(\mathcal{M})$. This simple observation can prove essential when establishing these conditions, as the water-filling surrogate costs of Markov chains have a much more intuitive structure (Lemma 4.2) than the surrogate costs of MDPs (Lemma 4.10). For example, while the surrogate costs of the MDPs corresponding to Additive PBs (Section 7) are very complicated, the surrogate costs of committing policies for the same setting admit an exceptionally simple form that can be exploited to prove a pointwise approximation.

Beyond Local Approximation. For one of our applications, namely the maximization version of Pandora's box with optional inspection (Section 9), local approximation turns out to be too stringent of a condition. As such, we introduce a slight weakening of local approximation, which we call *semilocal approximation* (Definition 9.4), that is tailored to this application. Roughly speaking, local approximation requires purely multiplicative suboptimality, while semilocal approximation allows for some additive suboptimality, too. See Section 9 for details.

[7] Appendix C exhibits an explicit example of this exponential scaling.

6 Pandora's Box with Partial Inspection (Minimization)

In classical Pandora's Box, the algorithm pays some cost c^o to open a box and observe its value, and is then allowed to select the box. We will now study the minimization version of the generalization called Pandora's Box with Partial Inspection (henceforth, PBPI) where the algorithm can additionally "peek" into the box at a smaller cost $c^p < c^o$ and learn its value. If upon obtaining this information, the algorithm wants to select the box, it must still open the box at a cost of c^o before accepting it. This presents a choice: in some cases it may be better to open the box outright, while in others it is better to pay the smaller peeking cost to potentially avoid paying the opening cost later.

Formally, we consider an instance with n partial inspection boxes (henceforth, PI-boxes) $\{\mathcal{B}_i\}_{i=1}^n$, where each box $\mathcal{B}_i = (\mathcal{D}_i, c_i^o, c_i^p)$ is characterized by a distribution $\mathcal{D}_i$ over values; an opening cost $c_i^o > 0$; and a peeking cost $c_i^p \in (0, c_i^o]$[8]. Each PI box $\mathcal{B}_i$ can be expressed as a Costly Information MDP with two possible actions, peeking and opening, where we can interpret the accept action after peeking as incurring a cost of c^o. An instance of PBPI corresponds to an instance of min-CICS over the corresponding MDPs. A pictorial representation of the different states of a box and the underlying MDP is shown in Figure 5.

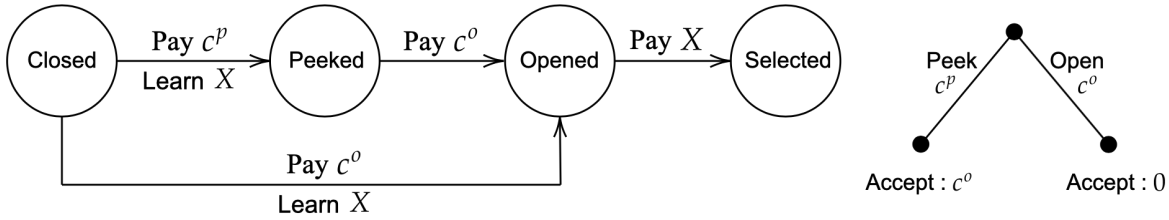


Figure 5: In PBPI, a peeking box $\mathcal{B} = (\mathcal{D}, c^o, c^p)$ is initially closed. In order to learn the value realization realization $X \sim \mathcal{D}$, the decision maker can either peek into the box (at a cost of c^p) or open it (at a cost of c^o). To accept the box, the decision maker must first open it and then pay its (now known) value X.

A committing policy for PBPI needs to decide in advance (perhaps randomly) whether each box $\mathcal{B}_i$ will be directly opened or peeked into and then (potentially) opened. Our main result in this section is the following:

THEOREM 6.1. *The commitment gap of matroid-PBPI is at most* $\sqrt{2}$.

In Section 6.1 we show a lower bound of 16/15 for the commitment gap of matroid-PBPI (Example 6.2) and a trivial upper bound of $\phi \approx 1.618$ (Lemma 6.3). We then prove Theorem 6.1 by showing that each PI box $\mathcal{B}$ admits a $\sqrt{2} \approx 1.414$ local approximation (Lemma 6.4) and employing Theorem 5.2; we note that the commitment achieving this guarantee decides whether each box will be directly opened or peeked before opened *deterministically*, based on the parameters of the box, and can be efficiently computed.

6.1 Lower and Upper Bounds on the Commitment Gap of PBPI. We begin by showing that that there are simple instances where all committing policies are sub-optimal up to a constant factor.

EXAMPLE 6.2. *Consider an instance of single-selection PBPI over two boxes. The first box has an opening cost of* 1*, a peeking cost of* $\frac{1}{4}$ *and its random value is* 0 *with probability* $\frac{1}{2}$ *and* 2 *otherwise. The second box has both an opening and peeking cost of* 0*, and its random value is* 2 *with probability* $\frac{1}{2}$ *and* ∞ *otherwise.*

- *Since opening the second box is free, we can assume without loss that all policies start by opening it and observing its value, call it* y*. The optimal policy will simply open and accept the first box if* $y = \infty$*. But if* $y = 2$*, it can peek into the first box and only open it if it contains a value of* 0*. The expected cost of this algorithm is* 15/8*.*
- *Now consider any policy that commits to opening or peeking into box* 1 before *observing* y*. If it commits to opening the first box, it accepts the value of this box regardless of* y*, as box* 1 *always has a value smaller than*

[8]Since we can only select an opened box, if $c_i^p \geq c_i^o$, then we can safely exclude the peeking action and the box reduces to a classical Pandora's Box.

y. The expected cost of this policy is 2. If it commits to peeking into the first box, then at $y = \infty$ it incurs a cost of 9/4 due to having to pay the extra peeking cost; and at $y = 2$ it incurs an expected cost of 7/4 by opening the first box only if it contains a value of 0. It's net cost is again 2.

Since both deterministic committing policies have an expected cost of 2, so does any randomized commitment. Consequently, the expected cost of the optimal adaptive policy is strictly smaller than that of the optimal committing policy.

Example 6.2 illustrates that the commitment gap of matroid-PBPI at least (16/15). On the positive side, it is easy to obtain an upper bound of 2. In particular, consider the policy that commits to peeking all boxes. This policy can mimic the optimal one as follows. Whenever the optimal algorithm peeks, so does this committing policy. Whenever the optimal algorithm opens without peeking, the committing policy peeks and then opens; on a box with costs c_i^p and c_i^o, this policy pays $c_i^p + c_i^o$ or at most twice the amount c_i^o paid by the optimal algorithm. In fact, we can further refine this argument by choosing the action we commit to more carefully: there always exists a simple commitment under which we can achieve a $\phi \approx 1.618$ approximation to the optimal adaptive policy.

LEMMA 6.3. *Consider any instance of matroid-PBPI and partition the n boxes into two sets*

$$O := \left\{ i \in [n] : \frac{c_i^o}{c_i^p} \leq 1 + \frac{c_i^p}{c_i^o} \right\}$$

and $P = [n] \setminus O$. The policy that commits to directly opening the boxes in O and peeking before opening the boxes in P achieves a ϕ-approximation to the optimal (non-committing) policy.

Proof. The proof relies on the crucial fact that opening and peeking into a PI box reveals precisely the same information to the decision maker; in particular, the value of the box. We construct a policy that commits to directly opening boxes $i \in O$ and peeking before opening boxes $i \in P$, while simultaneously mimicking the optimal adaptive policy. In particular, fix any box i.

- If the optimal never interacts with box i, neither does our policy.
- If the optimal directly opens box i, then (i) if $i \in O$ our policy also opens it and never peeks into it, and (ii) if $i \in P$ our policy first peeks into the box and then immediately opens it.
- If the optimal first peeks into box i, then (i) if $i \in P$ our policy also peeks into it (and then opens it whenever the optimal decides to open it), and (ii) if $i \in O$ our policy directly opens it and never peeks into it.
- If the optimal selects box i, so does our policy.

Observe that at any point the optimal and our policy have the exact same information, so we can keep mimicking the optimal decision tree. Furthermore, our policy clearly respects the given commitment. Finally, whenever the optimal selects a box, we can do the same as the set of our policy's opened boxes is always a superset of the optimal's opened boxes. This ensures feasibility of our algorithm under any combinatorial constraint $\mathcal{F}$.

Thus, the only difference between the costs of our policy and the optimal is due to differences in the selected actions. In particular, if $i \in P$ then if the optimal peeks into or ignores the box then so does the algorithm, with the worst case being the optimal directly opening the box. In that case, the optimal pays c_i^o whereas our policy pays $c_i^p + c_i^o$. On the other hand, if $i \in O$, then the worst case is if the optimal peeks into the box and decides not to open it; in that case the optimal pays c_i^p whereas our algorithm pays c_i^o. Finally, our algorithm pays precisely the same cost as the optimal for accepting boxes.

Combining everything, we conclude that our policy achieves an

$$\alpha := \max\left(\max_{i \in O}\left(\frac{c_i^o}{c_i^p}\right) , \max_{i \in P}\left(\frac{c_i^p + c_i^o}{c_i^o}\right) \right)$$

approximation to the optimal adaptive policy. For each box $i \in [n]$, let $\lambda_i = c_i^p / c_i^o \in (0, 1)$ and observe that by definition of the partition sets we have that $1/\lambda_i \leq 1 + \lambda_i$ if and only if $i \in O$. Thus, we obtain that

$$\alpha \leq \max_i \min(1 + \lambda_i, \frac{1}{\lambda_i}) \leq \max_{x \in (0,1)} \min(1 + x, \frac{1}{x}) = \phi.$$

as desired. ∎

We note that by using a *global argument* such as the one above, i.e. an argument where we charge each action of the committing policy directly to the optimal adaptive policy, one cannot improve on this upper bound of ϕ (for example, consider a setting where all the boxes have $c^p = 1$ and $c^o = \phi$). In order to achieve a better guarantee, we would need to leverage our knowledge of the value distributions. In the next section, we achieve this by establishing local approximation guarantees for PBPI.

6.2 Local Approximation Guarantees for PBPI. Given a PI box $\mathcal{B} = (\mathcal{D}, c^o, c^p)$, let g^p denote the water filling (a.k.a. Gittins) index of the policy that commits to peeking. Equivalently, g^p is the solution to the equation $c^p = \mathrm{E}_{X\sim\mathcal{D}}\left[(g^p - X - c^o)^+\right]$. We call g^p the **peeking index** of the box.

LEMMA 6.4. *Let $\mathcal{B} = (\mathcal{D}, c^o, c^p)$ be a PI box with peeking index g^p. Let π commit to the opening action whenever*

$$\frac{c^o}{c^p} \cdot \left(1 - \frac{c^o}{g^p}\right) \leq 1 + \min\left(\frac{c^p}{c^o}, \frac{c^o}{g^p}\right)$$

and to the peeking action otherwise. Then, π is a $\sqrt{2}$-local approximation to $\mathcal{B}$.

The rest of this section is devoted to proving Lemma 6.4. We first note the following characterization:

DEFINITION 6.5. (OPTIMALITY CURVES FOR PBPI) *The optimality curve of a PI box $\mathcal{B} = (\mathcal{D}, c^o, c^p)$ for cost minimization is given by*

$$f_{\mathcal{B}}(y) := \min\{y, f^o_{\mathcal{B}}(y), f^p_{\mathcal{B}}(y)\}$$

*where $f^o_{\mathcal{B}}(y) := c^o + \mathrm{E}_{X\sim\mathcal{D}}\left[\min\{y, X\}\right]$ is the optimality curve of the policy that commits to opening the box and $f^p_{\mathcal{B}}(y) := c^p + \mathrm{E}_{X\sim\mathcal{D}}\left[\min\{y, X + c^o\}\right]$ is the optimality curve of the policy that commits to peeking. We also define the **opening index** of the box, g^o, as the water filling index of the opening policy, equivalently, the solution to the equation $c^o = \mathrm{E}_{X\sim\mathcal{D}}\left[(g^o - X)^+\right]$. We depict these curves and indices in Figure 6.*

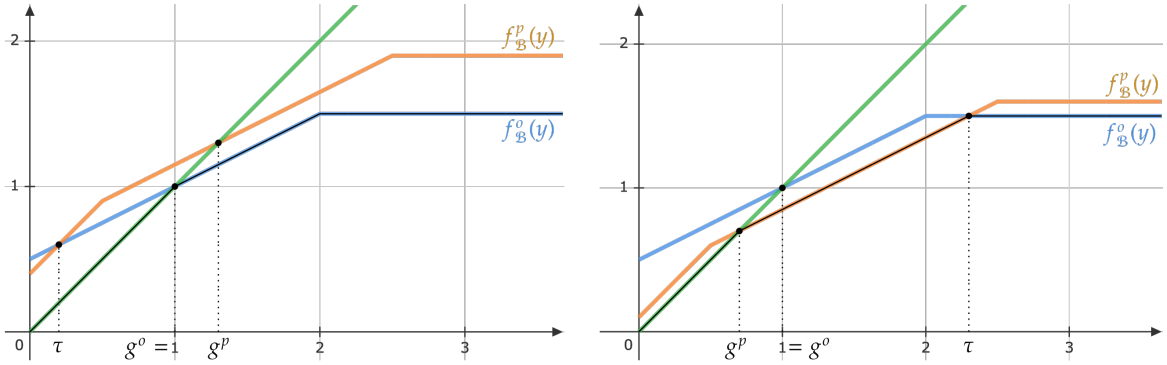


Figure 6: The optimality curves for a box $\mathcal{B}$ with opening cost $c^o = 0.5$, value $X = 0$ with probability 0.5 and $X = 2$ with probability 0.5 and peeking cost $c^p = 0.4$ (left) and $c^p = 0.1$ (right). The optimality curve of $\mathcal{B}$ is given by the minimum of the three curves. Observe that the curves of the opening and peeking action intersect at a unique point τ and that unless $g^p < g^o$, the opening action dominates the peeking action.

Observe that if the optimal adaptive policy prefers opening over peeking the box for some outside option y, then the same will be true for any $y' > y$. Also, for $y = 0$ peeking is clearly preferable to opening. Thus, the optimality curves $f^o_{\mathcal{B}}(y)$ and $f^p_{\mathcal{B}}(y)$ will have a unique intersection (up to an interval). We use τ to denote this intersection; formally, τ is the maximal solution to equation

$$\mathrm{E}_{X\sim\mathcal{D}}\left[(\tau - X)^+\right] - \mathrm{E}_{X\sim\mathcal{D}}\left[(\tau - X - c^o)^+\right] = c^o - c^p.$$

This implies that unless $g^p < g^o < \tau$, the opening action will dominate the peeking action for all values of the outside option y for which accepting it isn't optimal; as a consequence, such instances trivially admit a 1-local approximation by committing to the opening action. From now on, we assume that $g^p < g^o < \tau$. Lemma 6.4 is a consequence of the following claims, establishing local approximation guarantees for the opening and peeking actions respectively.

CLAIM 6.6. *For all $y \in \mathbb{R}$, it holds that* $\min\{y, f^o_{\mathcal{B}}(y)\} \leq \alpha \cdot f_{\mathcal{B}}(\frac{y}{\alpha})$ *for* $\alpha = \frac{c^o}{c^p} \cdot \left(1 - \frac{c^o}{g^p}\right)$.

CLAIM 6.7. *For all $y \in \mathbb{R}$, it holds that* $\min\{y, f^p_{\mathcal{B}}(y)\} \leq \alpha \cdot f_{\mathcal{B}}(\frac{y}{\alpha})$ *for* $\alpha = 1 + \min\left(\frac{c^p}{c^o}, \frac{c^o}{g^p}\right)$.

Note that the above claims immediately imply an α-local approximation with

$$\alpha = \min\left\{\frac{c^o}{c^p} \cdot \left(1 - \frac{c^o}{g^p}\right) \ , \ 1 + \frac{c^p}{c^o} \ , \ 1 + \frac{c^o}{g^p}\right\}.$$

Fixing c^o and c^p and letting $\lambda := c^p/c^o \in (0,1)$, the minimum of the first and third terms is maximized at $g^p = c^o \cdot \frac{1+\lambda}{1-\lambda}$; so we have $\alpha \leq \min\{1+\lambda, \frac{2}{1+\lambda}\} \leq \sqrt{2}$. We conclude this section with the proofs of the two claims.

Proof of Claim 6.6. By concavity, we have that $f^o_{\mathcal{B}}(y) \leq \alpha f^o_{\mathcal{B}}(\frac{y}{\alpha})$ for all $y \in \mathbb{R}$ and $\alpha \geq 1$; thus, we only need to verify the condition for y such that $f_{\mathcal{B}}(\frac{y}{\alpha}) = f^p_{\mathcal{B}}(\frac{y}{\alpha})$. In other words, we only need to verify

$$\min\{y, f^o_{\mathcal{B}}(y)\} \leq \alpha f^p_{\mathcal{B}}(\frac{y}{\alpha}) \quad \forall y \in [\alpha g^p, \alpha\tau].$$

For start, we will show that the condition holds for $\alpha = g^o/g^p$; notice that for this parameter, $f^o_{\mathcal{B}}(y) \leq y$ for all $y \in [\alpha g^p, \alpha\tau]$ and thus we can now write our condition as

$$D_\alpha(y) := \alpha f^p_{\mathcal{B}}(\frac{y}{\alpha}) - f^o_{\mathcal{B}}(y) \geq 0 \quad \forall y \in [\alpha g^p, \alpha\tau].$$

Let $F(\cdot)$ denote the CDF of distribution $\mathcal{D}$. By definition of $f^o_{\mathcal{B}}(y)$ and $f^p_{\mathcal{B}}(y)$, we have

$$\frac{d}{dy} f^o_{\mathcal{B}}(y) = 1 - F(y) \ \ , \ \ \frac{d}{dy} f^p_{\mathcal{B}}(y) = 1 - F(y - c^o)$$

and by taking the derivative of $D_\alpha(\cdot)$, we immediately obtain that $D_\alpha(\cdot)$ is a weakly increasing function of y and thus the condition only needs to hold on $y = \alpha g^p$. Observe that for $\alpha = g^o/g^p$ and $y = \alpha g^p = g^o$, we have

$$D_\alpha(y) = \frac{g^o}{g^p} \cdot f^p_{\mathcal{B}}(g^p) - f^o_{\mathcal{B}}(g^o) = g^o - g^o = 0$$

and thus we obtain that the opening action always achieves a g^o/g^p local approximation.

The final step of the proof will be to show that

$$\frac{g^o}{g^p} \leq \frac{c^o}{c^p} \cdot \left(1 - \frac{c^o}{g^p}\right)$$

For this purpose, we define the function $h(z) := \mathrm{E}_{X \sim \mathcal{D}}\left[(z - X)^+\right]$ and observe that $h'(z) = F(z)$; thus, $h(z)$ is an increasing and convex function of z with $h(0) = 0$. Furthermore, by definition of the indices and the curves we have that $h(g^o) = c^o$ and $h(g^p - c^o) = c^p$. Thus, by convexity, we immediately obtain that

$$\frac{h(g^o)}{g^o} \geq \frac{h(g^p - c^o)}{g^p - c^o} \Rightarrow \frac{c^o}{g^o} \geq \frac{c^p}{g^p - c^o}$$

from which the claim follows. ∎

Proof of Claim 6.7. By concavity, we have that $f^p_{\mathcal{B}}(y) \leq \alpha f^p_{\mathcal{B}}(\frac{y}{\alpha})$ for all $y \in \mathbb{R}$ and $\alpha \geq 1$; thus, we only need to verify the condition for y such that $f_{\mathcal{B}}(\frac{y}{\alpha}) = f^o_{\mathcal{B}}(\frac{y}{\alpha})$; notice that this corresponds to $y \geq \alpha\tau > g^p$ and thus $f^p_{\mathcal{B}}(y) \leq y$. In other words, we only need to verify

$$f^p_{\mathcal{B}}(y) \leq \alpha f^o_{\mathcal{B}}(\frac{y}{\alpha}) \quad \forall y \geq \alpha\tau.$$

Let $D_\alpha(y) := \alpha \cdot f^o_{\mathcal{B}}(y/\alpha) - f^p_{\mathcal{B}}(y)$. We need $D_\alpha(y) \geq 0$ for $y \geq \alpha\tau$. The condition immediately holds at $y = \alpha\tau$ since

$$\alpha f^o_{\mathcal{B}}(\frac{\alpha\tau}{\alpha}) = \alpha f^o_{\mathcal{B}}(\tau) = \alpha f^p_{\mathcal{B}}(\tau) \geq f^p_{\mathcal{B}}(\alpha\tau)$$

by concavity of $f^p_{\mathcal{B}}(\cdot)$. Next, observe that

$$\frac{d}{dy}D_\alpha(y) = \frac{d}{dy}f^o_{\mathcal{B}}(\frac{y}{\alpha}) - \frac{d}{dy}f^p_{\mathcal{B}}(y) = F(y - c^o) - F(\frac{y}{\alpha})$$

and thus $D_a(y)$ gets minimized at $y = \frac{\alpha \cdot c^o}{\alpha - 1}$. If $\frac{\alpha \cdot c^o}{\alpha - 1} \leq \alpha\tau$ or equivalently $\alpha \geq 1 + c^o/\tau$, then $\frac{d}{dy}D_\alpha(y) \geq 0$ in the area of interest; thus, we obtain that the peeking action always achieves a $1 + \frac{c^o}{\tau} \leq 1 + \frac{c^o}{g^p}$ local approximation. To complete the proof, we need to show that the peeking action also achieves a $1 + \frac{c^p}{c^o}$ local approximation or equivalently that for $\alpha = 1 + \frac{c^p}{c^o}$ we have that $\min_{y \geq \alpha\tau} D_\alpha(y) = D(\frac{\alpha \cdot c^o}{\alpha - 1}) \geq 0$.

Once again, we consider the function $h(z) = \mathrm{E}_{X \sim \mathcal{D}}\left[(z - X)^+\right]$; this time, we observe that by definition we have

$$f^o_{\mathcal{B}}(y) = c^o + y - h(y) \ \ , \ \ f^p_{\mathcal{B}}(y) = c^p + y - h(y - c^o).$$

From this, the condition $D(\frac{\alpha \cdot c^o}{\alpha - 1}) \geq 0$ translates to

$$h(\frac{c^o}{\alpha - 1}) \leq \frac{\alpha \cdot c^o - c^p}{\alpha - 1}$$

and for $\alpha = 1 + \frac{c^p}{c^o}$, this statement is equivalent to

$$h(\frac{c^o \cdot c^o}{c^p}) \leq \frac{c^o \cdot c^o}{c^p}$$

which is clearly true, by definition of $h(\cdot)$. □

7 Additive Pandora's Box (Minimization)

In this section, we study a different generalization of Pandora's Box which we call *Additive Pandora's Box* (henceforth, APB). In this setting, the value of each (additive) box is given by the sum of independent random variables, all of which need to be separately probed by paying a corresponding probing cost. The objective is to minimize the total sum of opening costs plus the additive value of the accepted box.

The primary motivation for studying APB is that it constitutes a special case of the notoriously challenging *Pandora's Shortest Path* problem, where the edge weights of a given graph correspond to independent Pandora's boxes and the objective is to accept a set of boxes that form a path between two given vertices s, t. Due to the fact that path constraints do not admit frugal algorithms, no results are currently known for this problem and it is left as an open direction by [Singla, 2017]. It is not hard to see that APB corresponds to the special case of Pandora's Shortest Path where the graph consists of parallel paths between s and t; each path is then modeled via an additive Pandora's Box (the components of which correspond to the path's edges) and the path constraint translates to a single-selection constraint over the additive boxes.

Formally, an instance of APB consists of n-additive boxes $\{\mathcal{B}_i\}_{i=1}^n$. Each box $\mathcal{B}_i = (\vec{\mathcal{D}}_i, \vec{c}_i)$ is characterized by $k_i \geq 1$ random variables, each distributed independently according to $\mathcal{D}_{ij}$ and with an opening cost of $c_{ij} \geq 0$ for $j \in [k_i]$. The algorithm can observe the realization $V_{ij} \sim \mathcal{D}_{ij}$ by paying c_{ij} and can accept the box $\mathcal{B}_i$ only after all the k_i random variables have been observed, at a cost of $\sum_{j=1}^{k_i} V_{ij}$. At any point, the algorithm can adaptively select which additive box to explore, as well as which of its components to probe.

We can immediately cast this problem as a min-CICS instance where each additive box corresponds to an MDP $\mathcal{M}$ (which we call an additive-MDP), capturing all the different (adaptive) probing orders for the components of the box. A commitment for an additive-MDP corresponds to a protocol that determines which component to probe first, which second depending on the realization of the first value etc. Our main contribution is a to show that the commitment gap of APB is constant, and does not depend on the number of components.

THEOREM 7.1. *The commitment gap of matroid-APB is at most* 2.

Our proof of Theorem 7.1 is existential and does not indicate how to efficiently construct the committing policy that achieves this bound. One could consider a restricted class of committing policies, which we call **static commitment policies**, where the components of each additive box are always probed in a fixed order. Trivial arguments suffice to show that static policies probing the components in increasing order of costs achieve a commitment gap of k. By considering the structure of water filling surrogate costs, we can obtain an improved bound on the gap:

THEOREM 7.2. *Let $\mathcal{M}$ be any additive-MDP. For $k = 2$, there exists a static commitment policy that achieves a $\phi \approx 1.618$-pointwise approximation. For $k \geq 3$, the static policy of minimum index achieves a $\left(1+\lfloor \frac{k+1}{2} \rfloor\right)$-pointwise approximation.*

In order to constructively achieve the approximation factor guaranteed by Theorem 7.2, one can compute the indices of the $k!$ static policies for each constituent MDP and commit to the one with the smallest index. We leave open the questions of whether it is possible to achieve the same approximation more efficiently, and whether one could efficiently compute a committing policy that admits a constant factor gap. The rest of this section is dedicated to proving Theorems 7.1 and 7.2.

Proof of Theorem 7.1. Our proof boils down to proving that each additive-MDP admits a 2-pointwise approximation under a suitably chosen commitment and then employing Fact 5.4 and Theorem 5.2. Our proof consists of two steps: first, we identify a simple structure for the surrogate costs under any commitment for APB and translate pointwise-approximation into a simple condition for the commitment. Then, we show that there exists some commitment satisfying this condition.

Step 1. Let $\mathcal{M}$ be any additive-MDP with $k \geq 1$ components corresponding to values $V_i \sim \mathcal{D}_i$ and costs $c_i \geq 0$ for $i \in [k]$. Let $\pi \in C(\mathcal{M})$ be any commitment; observe that π specifies which value to be probed first, which to be probed second depending on the first's realization and so on until all k-values have been probed. Therefore, $\mathcal{M}^\pi$ is a tree-structured Markov chain of height k, the terminal states $t(\vec{V})$ of which correspond to all the possible realizations of the value vector $\vec{V} = (V_1, \cdots, V_k)$. Furthermore, the probability of reaching such a terminal state $t(\vec{V})$ is always equal to the probability of $\vec{V}$ being realized, independently of π's probing protocol. Therefore, the water filling surrogate cost $W^*_{\mathcal{M}^\pi}$ for any commitment π is obtained by sampling $\vec{V} \sim \vec{\mathcal{D}}$ and then returning a (commitment-specific) surrogate cost $\rho^\pi(t(\vec{V}))$. Due to this structure, the α-pointwise approximation condition for additive-MDPs translates to proving that there exists some commitment $\pi \in C(\mathcal{M})$ such that

$$\rho^\pi(t(\vec{V})) \leq \alpha \cdot \rho^{\pi'}(t(\vec{V}))$$

for all vectors $\vec{V} \sim \vec{\mathcal{D}}$ and all *deterministic* commitments $\pi' \in C(\mathcal{M})$.

Up next, we will use the inductive definition of water filling amortization to obtain structure on these surrogate costs. Fix a commitment $\pi \in C(\mathcal{M})$ and the underlying Markov chain $\mathcal{M}^\pi$. For each state s of $\mathcal{M}^\pi$, we define the following quantities:

1. The **index** $g^\pi(s)$ of state s is defined as the water filling index of the sub-chain rooted at s, if all values that have been already probed have been realized to 0. In other words, $g^\pi(s)$ corresponds to the index of the committing policy that proceeds as π after state s is reached, without the additive cost of already observed values. Clearly, $g^\pi(t(\vec{V})) = 0$ for all terminal states $t(\vec{V})$ and $g^\pi(r) = g^\pi$ for the root r of $\mathcal{M}^\pi$.

2. The **value** $V^\pi(s)$ of state s is defined as the sum of all the realized values in the trajectory from the root to s; clearly, we have that $V^\pi(r) = 0$ and $V^\pi(t(\vec{V})) = \sum_{i=1}^k V_i$ for all terminal states $t_{\vec{V}}$.

The key observation is that when amortizing some state s, all the terminal states in it's subtree will share the same additive offset $V^\pi(s)$, corresponding to the sum of realized values leading to s. It is immediate by the definition of the water filling amortization that such an additive offset does not affect the cost shares of vertex s. Also, recall that water filling amortization is performed in a bottom-up manner. Thus, the surrogate cost of a terminal state $t(\vec{V})$ immediately after one of its ancestors s gets amortized will be precisely $V^\pi(s) + \max(g^\pi(s), \rho'(t(\vec{V})))$, where $\rho'(t(\vec{V}))$ is the surrogate cost of $t(\vec{v})$ immediately after it's ancestor that is a child of s was amortized. Performing this step for all ancestors of $t(\vec{V})$, we obtain that

$$\rho^\pi(t(\vec{V})) = \max_{i=1}^{k+1} \left(V^\pi(s_i) + g^\pi(s_i) \right)$$

where $(r = s_1, s_2, \cdots, s_k, s_{k+1} = t(\vec{V}))$ is the unique path in $\mathcal{M}^\pi$ that reaches $t(\vec{V})$.

From this expression, it becomes clear that $\rho^\pi(t(\vec{V})) \geq \max(g^\pi, \sum_{i=1}^k V_i)$. Furthermore, if it is the case that $g^\pi \geq g^\pi(s)$ for all states s, then we can easily upper bound $\rho^\pi(t(\vec{V})) \leq g^\pi + \sum_{i=1}^k V_i$. We refer to commitments

that have this property as **root dominant**. Combining everything, we then obtain that for any root dominant commitment π, any commitment π' and any terminal state $t(\vec{V})$, we have that

$$\frac{\rho^{\pi}(t(\vec{V}))}{\rho^{\pi'}(t(\vec{V}))} \leq \frac{g^{\pi} + \sum_{i=1}^{k} V_i}{\max(g^{\pi'}, \sum_{i=1}^{k} V_i)} \leq 1 + \frac{g^{\pi}}{g^{\pi'}}.$$

In other words, we have shown that any root dominant commitment π will achieve an α-pointwise approximation for $\alpha = 1 + g^{\pi}/\min_{\pi' \in C(\mathcal{M})} g^{\pi'}$. The proof of Theorem 7.1 is then completed by arguing that there exists a root dominant commitment π that has minimum index across all committing policies.

Step 2. Let π be any commitment of minimum index. If π is root dominant, then we are obviously done. Otherwise, there exists at least one state s_1 in $\mathcal{M}^{\pi}$ such that $g^{\pi}(s_1) > g^{\pi}$. Among all these states, we consider one of minimum distance to the root. Notice that s_1 must have at least one sibling state s_2 with $g^{\pi}(s_2) \neq g^{\pi}(s_1)$; otherwise, the index of of the parent state of s_1 will be at least $g^{\pi}(s_1)$ (by the inductive definition of indices) and thus s_1 won't be a maximum height state satisfying the condition.

Let π' be the commitment that is obtained by substituting the sub-tree rooted at s_1 in $\mathcal{M}^{\pi}$ with the sub-tree rooted at s_2; notice that both of these subtrees correspond to commitments over the (same) set of unprobed values and thus such a substitution is always allowed (i.e. $\pi' \in C(\mathcal{M})$ as well). Up next, we will show that $g^{\pi} \geq g^{\pi'}$; then, π' will also be a minimum index commitment and we can apply the same process to it. Furthermore, each substitution reduces the heteromorphity of the underlying chain (i.e. the number of different child sub-trees that a state can have) so our iterative process is guaranteed to terminate after some finite number of steps, resulting in a commitment that is both root dominant and has minimum index.

We finally argue that $g^{\pi} \geq g^{\pi'}$. Since g^{π} is the water filling index of $\mathcal{M}^{\pi}$, it is by definition at least as large as the minimum surrogate cost of *any* valid amortization of $\mathcal{M}^{\pi}$. Let $b = \{b_{st}\}_{s,t}$ denote the water filling cost shares of $\mathcal{M}^{\pi}$ and $b' = \{b'_{st}\}_{s,t}$ denote the water filling cost shares of $\mathcal{M}^{\pi'}$ Now, consider the following amortization $\hat{b} = \{\hat{b}_{st}\}_{s,t}$ of $\mathcal{M}^{\pi}$:

- For each state s that does not belong in the sub-tree rooted at s_1, let $\hat{b}_{st} = b'_{st}$ for all terminals t.
- For each state s in the sub-tree rooted at s_1 (including s_1), let $\hat{b}_{st} = b_{st}$ for all terminals t.

Since all the states of $\mathcal{M}^{\pi}$ and $\mathcal{M}^{\pi'}$ that are not ancestors of state s_1 have the same distribution over terminal states, this is indeed a valid amortization. Furthermore, any terminal state t that is not a descendant of s_1 will receive precisely the same cost shares as it does in $\mathcal{M}^{\pi'}$ and thus it's surrogate cost will be at least $g^{\pi'}$. On the other hand, every terminal state t that is a descendant of s_1 receives the same cost shares as it did in $\mathcal{M}^{\pi}$ up until state s_1 gets amortized; by definition, this implies that it's surrogate cost will be at least $g^{\pi}(s_1) + V^{\pi}(s_1) \geq g^{\pi}(s_1)$. Thus, we conclude that $g^{\pi} \geq \min(g^{\pi'}, g^{\pi}(s_1)) = g^{\pi'}$ as desired. □

Proof of Theorem 7.2. We will separately prove the theorem for $k = 2$ and $k \geq 3$.

The $k = 2$ case. We use (X, c_x) and (Z, c_z) to denote the $k = 2$ components of the additive-MDP and g_x, g_z to denote their respective water filling indices. Notice that there are only two (static) committing policies to consider: π_{XZ} and π_{ZX} depending on whether X is probed first or not. We use g_{xz} and g_{zx} for their respective indices. Using the same structural properties of surrogate costs for additive-MDPes as in the proof of Theorem 7.1, we have that π_{XZ} admits an α_{xz}-pointwise approximation for

$$\alpha_{xz} = \max_{t=(x,z)} \frac{\rho_{xz}(t)}{\rho_{zx}(t)} = \max_{t=(x,z)} \frac{\max(g_{xz}, g_x + z, z + x)}{\max(g_{zx}, g_z + x, z + x)} \leq \max_{t=(x,z)} \max(1, \frac{g_{xz}}{g_{zx}}, \frac{g_x + z}{\max(g_{zx}, g_z + x, z + x)})$$

and for the last term in the maximum, since $x \geq 0$, we have that

$$\alpha_{xz} \leq \max_{t=(x,z)} \max(1, \frac{g_{xz}}{g_{zx}}, \frac{g_x + z}{\max(g_{zx}, z)}) \leq \max(\frac{g_{xz}}{g_{zx}}, 1 + \frac{g_x}{g_{zx}})$$

and likewise we obtain that

$$\alpha_{zx} \leq \max(\frac{g_{zx}}{g_{xz}}, 1 + \frac{g_z}{g_{xz}}).$$

Finally, we will now prove that $\min(\alpha_{zx}, \alpha_{zx}) \leq \phi$ always. By re-labeling and re-scaling we can assume without loss that $g_{xz} \leq g_{zx} = 1$. Furthermore, we have that $g_{xz} \geq g_x + g_z$. To see why this is true, consider an amortization $\mathcal{M}^{\pi_{xz}}$ where each (height 1) Z-vertex sends the same cost shares b_z as in the water filling amortization of (Z, c_z) to any terminal (x, z), and the (unique) X-vertex sends the same cost shares b_x as in the water filling amortization of (X, c_x) to any terminal (x, z) - this amortization covers the opening costs by definition, and the surrogate cost of terminal (x, z) is $x+z+b_x+b_z = \rho_x(x)+\rho_z(z) \geq g_x+g_z$. Since water-filling amortization maximizes the minimum surrogate cost across all amortizations, we indeed have $g_{xz} \geq g_x + g_z$. Thus, we have $g_x + g_z \leq g_{xz} \leq g_{zx} = 1$. These in turn imply that $\alpha_{xz} \leq 1 + g_x$ and $\alpha_{zx} \leq \frac{1}{g_x+g_z} \max(1, g_x + 2g_z)$. The proof is then completed by showing

$$\alpha := \min\left(1 + g_x, \frac{1}{g_x + g_z} \max(1, g_x + 2g_z)\right) \leq \phi$$

for all $g_x, g_z \geq 0$ with $g_x + g_z \leq 1$. By fixing g_x, it is not hard to see that the second term gets maximized at either $g_z = 0$ or $g_z = 1 - g_x$. Thus, we have

$$\alpha \leq \max_{x \in [0,1]} \min\left(x + 1, \max(\frac{1}{x}, 2 - x)\right) = \phi.$$

The $k \geq 3$ case. Let $\pi \in \mathcal{C}(\mathcal{M})$ be any commitment of index g^π for an additive-MDP $\mathcal{M}$ over k-components and let $\alpha = \lfloor \frac{k+1}{2} \rfloor$. We will show that there exists a static committing policy π' such that $g^{\pi'} \leq \alpha \cdot g^\pi$. Then, by instantiating π as any minimum index policy and observing that all static policies are by definition root dominant (i.e. the root state has maximum index across all states), step 1 in the proof of Theorem 7.1 immediately gives us the proof of Theorem 7.2.

Consider the Markov chain $\mathcal{M}^\pi$. Like in the proof of Theorem 7.1, we define the index $g^\pi(s)$ of a state s as the water filling index of the sub-chain rooted at s if all previously probed values are realized at 0, and its value $V^\pi(s)$ as the sum of all realized values in the trajectory from the root to s. Observe that these indices and values uniquely determine the water filling cost shares that each state s sends to its downwards terminal states in $\mathcal{M}^\pi$; in particular, we have that

$$\begin{aligned} b_{st} &= \max\left(0, g^\pi(s) + V^\pi(s) - \max_{s' \in P(s,t)} (g^\pi(s') + V^\pi(s'))\right) \\ &= \max\left(0, g^\pi(s) - \max_{s' \in P(s,t)} (g^\pi(s') + V^\pi(s, s'))\right) \end{aligned}$$

where the inner maximum is taken over all states in the (unique) path from s to t, excluding state s, and $V^\pi(s, s')$ denotes the sum of all realizations in the trajectory from s to s'. The above equality is a direct consequence of the fact that the minimum surrogate cost in the sub-tree of state s immediately after s gets amortized is by definition $g^\pi(s) + V^\pi(s)$.

We will now define the static policy π'. Observe that any trajectory from the root of $\mathcal{M}^\pi$ to a terminal state defines a fixed ordering of the k-components; we define π' as the ordering corresponding to a trajectory of lexicographically maximum index. In other words, π' probes the same first box as π, then the second box corresponding to a child of the root with maximum index etc. By re-labeling, we assume that π' is the static ordering $k \mapsto (k-1) \mapsto \cdots \mapsto 1$.

We now turn our attention to the Markov chain $\mathcal{M}^{\pi'}$. We use $h(s)$ to denote the height of state s; this is 0 for terminal states and k for the root. Notice that in $\mathcal{M}^{\pi'}$, all the states s with $h(s) = j \in [k]$ correspond to the same (remaining) fixed probing order $j \mapsto (j-1) \mapsto \cdots \mapsto 1$, and will have the same index $g^{\pi'}(s) = g(j)$, with $g^{\pi'} = g(k)$. Furthermore, each state s of $\mathcal{M}^{\pi'}$ at height $j \in [k]$ can be mapped to the unique height j state of $\mathcal{M}^\pi$ that was chosen by our process; we use s_j to denote this state.

We will prove that for all $j \in [k]$, it holds that $g(j) \leq \alpha(j) \cdot g^\pi(s_j)$ for $\alpha(j) := \lfloor \frac{j+1}{2} \rfloor$. This would directly imply that $g^\pi = g(k) \leq \alpha(k) \cdot g^\pi(s_k) = \alpha(k) \cdot g^\pi$, completing the proof. In order to upper bound the indices $g(j)$ of $\mathcal{M}^{\pi'}$, it suffices to argue that the suggested upper bounds generate enough cost shares to cover the amortization of the costs. Since each state s of $\mathcal{M}^{\pi'}$ at height $j \in [k]$ shares the same distribution over terminal states with the state s_j of $\mathcal{M}^\pi$, it will suffice to point-wise compare these cost shares. In other words, it suffices to prove that

$$\max\left(0, \alpha(j) \cdot g^\pi(s_j) - \max_{s' \in P'(s,t)} (\alpha(h(s')) \cdot g^\pi(s_{h(s')}) + V^{\pi'}(s, s'))\right) \geq \max\left(0, g^\pi(s_j) - \max_{s' \in P(s_j,t)} (g^\pi(s') + V^\pi(s_j, s'))\right)$$

for all $j \in [k]$. Our proof will be completed by arguing that our specific definition of $\alpha(j)$ satisfies

$$\max_{s' \in P'(s,t)} \left(\alpha(h(s')) \cdot g^{\pi}(s_{h(s')}) + V^{\pi'}(s, s') \right) \leq \alpha(j) \cdot \max_{s' \in P(s_j,t)} \left(g^{\pi}(s') + V^{\pi}(s_j, s') \right)$$

for all $j \in [k]$, all states s with $h(s) = j$ and all terminal states t.

The maximum in the left hand side will get realized at some $s' \in P'(s,t)$ with $h(s') := j' \leq j-1$ (as $h(s) = j$). If $j' < j-1$, then we can upper bound the left hand side by $a(j') \cdot g^{\pi}(s_{j'}) + V^{\pi'}(s,t)$. Since π' is a static policy, it is not hard to see that the indices in any trajectory from the root to a terminal will form a decreasing sequence, so $g^{\pi}(s_{j'}) \leq g^{\pi}(s_j)$. Furthermore, we have that $V^{\pi'}(s,t) = V^{\pi}(s_j, t)$ and thus our inequality holds as long as $\alpha(j) \geq \alpha(j') + 1$ which is always the case for $j' < j-1$.

It remains to argue about $j' = j-1$. In that case, the left hand side is $\alpha(j-1) \cdot g^{\pi}(s_{j-1}) + V_j$ for some realization V_j of the j-th component. We will upper bound this through the first term in the right hand side maximum, corresponding to some child vertex $\hat{s}$ of s_j; in that case, we need to show that

$$\alpha(j-1) \cdot g^{\pi}(s_{j-1}) + V_j \leq \alpha(j) \cdot g^{\pi}(\hat{s}) + V_j$$

and the proof is completed by the fact that our policy selected a trajectory of lexicographically maximum indices and thus $g^{\pi}(s_{j-1}) \leq g^{\pi}(s')$ as s_{j-1} and s' are both children of the same state s_j in $\mathcal{M}^{\pi}$. ∎

8 The Weighing Scale Problem (Minimization)

We will now introduce the *Weighing Scale* (henceforth, WS) problem. A decision maker is presented with n alternatives (X_i, c_i) and a combinatorial constraint $\mathcal{F} \subseteq 2^{[n]}$; $X_i \geq 0$ is the random value of the alternative realized independently by a known distribution and $c_i \geq 0$ is a weighing cost. The only way the decision maker can determine any further information about each value X_i beyond its distribution, is to use a weighing scale to compare it against some fixed threshold t of their choosing at the additional cost of c_i and learn whether $X_i \leq t$ or not. The process terminates with the decision maker selecting a feasible set of alternatives $S \subseteq \mathcal{F}$, paying their total value. Note that each alternative (X_i, c_i) corresponds to a Costly Information MDP $\mathcal{M}_i$ that captures the information acquisition process described above. In particular, the available actions at the starting state of each MDP $\mathcal{M}_i$ are as follows:

1. Pick a threshold $t \in \text{support}(X_i)$ and weigh the alternative against it at a cost of c_i. Upon taking one of these actions, the MDP advances to one of two random subprocesses $\mathcal{M}_i^{\leq t}$ and $\mathcal{M}_i^{>t}$, defined over the random variables $(X_i | X_i \leq t)$ and $(X_i | X_i > t)$ respectively, based on the outcome of the weighing.
2. Commit no more weighings of the alternative; this is a 0-cost action resulting to a terminal state x_i of value $v(x_i) := \mathrm{E}[X_i]$.

From this equivalence, an instance of WS corresponds to an instance of min-CICS over the corresponding MDPs.[9] Our main result for this section is the following:

THEOREM 8.1. *The commitment gap of matroid-WS is at most $O(\max_i \kappa_i)$, with the parameter κ_i for each alternative $i \in [n]$ defined as*

$$\kappa_i := \frac{\mu_i}{M_i} + \log \frac{\mu_i}{g_i}$$

where $\mu_i = \mathrm{E}[X_i]$ is the expected value of X_i, M_i is the median value of X_i, and g_i denotes the Gittins index of the alternative; i.e. the solution to the equation $c_i = \mathrm{E}[(g_i - X_i)^+]$.

We prove Theorem 8.1 by showing that for each alternative $i \in [n]$, the MDP $\mathcal{M}_i$ admits an $O(\kappa_i)$-pointwise approximation. Furthermore, the commitment π that achieves this condition corresponds to a very natural and efficient algorithm which we call *One-Sided Halving*. Finally, we note that without any assumptions on the distributions of the values, the parameters κ_i can be unbounded; in particular, for heavy tail distributions, μ_i / M_i can be arbitrarily large. We show that this dependence is necessary for any pointwise approximation guarantee. Better bounds may be possible via the weaker notion of local approximation or a global argument.

[9] Technically, our framework does not capture infinite horizon MDPs. However, observe that whenever the decision maker has identified that the value of the alternative lies in some interval of length $\leq c$, performing any extra weighings is suboptimal. Thus, the corresponding MDPs for the WS problem are finite horizon without loss, even for continuous random variables X.

THEOREM 8.2. *For any constant $\alpha \geq 1$, there exists a WS alternative that does not admit an α-pointwise approximation.*

The rest of this section is dedicated towards proving Theorem 8.1 (Section 8.1) and Theorem 8.2 (Section 8.2).

8.1 Proving the Upper Bound. We begin by introducing the committing policy for WS described below, which we call the *One-Sided Halving algorithm*. The policy begins by weighing the alternative against some threshold t_2; if the alternative is larger, then the policy commits to no more weighings and if it is smaller, it halves its threshold from t_2 to $t_2/2$ and repeats, until the threshold reaches some fixed lower bound t_1.

ALGORITHM 8.3. (ONE-SIDED HALVING ALGORITHM.)

1: **Input.** An MDP $\mathcal{M}$ corresponding to a WS alternative (X, c) and two thresholds $0 < t_1 < t_2$.
2: Set $t = t_2$.
3: **while** $t \geq t_1$ **do**
4: Weigh the random variable X against t. **if** $X \leq t$ **then** t = t/2 **else** break.
5: Commit to no more weighings for the alternative.

Let h be the solution to equation $c = \mathrm{E}\left[(X-h)^+\right]$. The following lemma, combined with our reduction from pointwise to local approximation (Fact 5.4) and our local approximation composition result (Theorem 5.2) directly implies Theorem 8.1.

LEMMA 8.4. *Committing to no weighings if $g > \min\{\mu, M\}$, and otherwise committing to the One-Sided Halving algorithm with $t_1 = g$ and $t_2 = \min(M, h)$ achieves an $O(\kappa)$-pointwise approximation.*

Before proving Lemma 8.4, we provide some intuition on how this result is obtained. A commitment $\pi \in \mathcal{C}(\mathcal{M})$ for WS will declare in advance a decision tree over pre-specified thresholds against which it will weigh X, resulting in a Markov chain $\mathcal{M}^\pi$. Importantly, π will end up partitioning the support $\mathcal{X}$ of distribution $\mathcal{D}$ into a set of intervals, corresponding to its terminal states. Furthermore, the probability of running the Markov chain $\mathcal{M}^\pi$ and ending up in a terminal state t corresponding to some interval I will be precisely $\Pr[X \in I]$. A pictorial representation of such Markov chains is given in Figure 7. By directly relating the water filling amortized cost on these intervals for the One-Sided Halving algorithm to the surrogate cost of the underlying MDP, we are able to obtain our pointwise approximation guarantees.

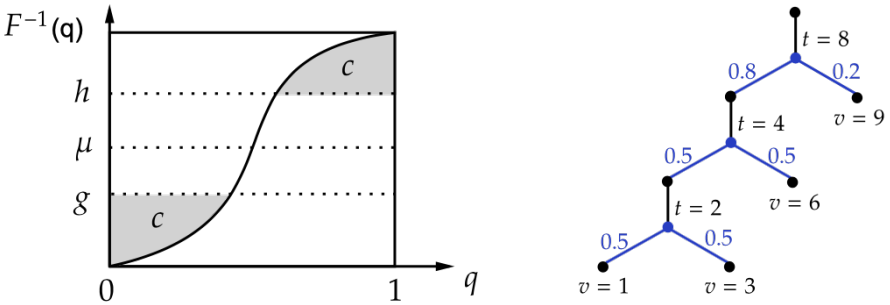


Figure 7: The indices g and h of the alternative can be obtained by the inverse CDF of the random value; the highlighted ares equal the weighing cost c. The figure on the right corresponds to the Markov chain for One-Sided Halving instantiated with threshold $t_1 = 2$, $t_2 = 8$ for an instance with value $X \sim \mathrm{Unif}[0, 10]$. The terminal states partition the support of X; the probability and value at each terminal equals the probability and the conditional expectation of the corresponding interval.

Proof of Lemma 8.4. We fix a WS alternative $\mathcal{M}$ corresponding to random value X of support $\mathcal{X} := \mathrm{support}(X)$ and weighing cost c. We use μ, M, g, h and κ to denote the parameters of the alternative, as previously defined. In order to develop our pointwise approximation guarantees, we first need to obtain expressions for the surrogate costs.

Surrogate cost of $\mathcal{M}$. We begin by noting that the optimality curve of $\mathcal{M}$ has a very simple form. Indeed, consider the local game $(\mathcal{M}, y)$; the optimal adaptive policy has only three choices available: either accept the outside option at a cost of y, or accept the alternative without performing any weighings at an expected cost of μ or perform a single weighing of the alternative against y to determine which of the two costs is smaller. Thus, we immediately obtain that $f_{\mathcal{M}}(y) = \min(y, \mu, c + \mathrm{E}[\min\{y, X\}])$.

Observe that by definition, g corresponds to the maximal threshold for which $y \leq c + \mathrm{E}[\min\{y, X\}]$ for all $y \leq g$. Thus, if $\mu \leq g$, then the cost $c + \mathrm{E}[\min\{y, X\}]$ will always be dominated by either y or μ, making the weighing action universally sub-optimal. If that's the case, then we can safely commit to blindly accepting the alternative without performing any weighings, achieving a 1-local approximation. Thus, from now on we will be assuming that $g < \mu$. Recall that g satisfies $c = \mathrm{E}[(g - X)^+]$; h satisfies $c = \mathrm{E}[(X - h)^+]$; and μ satisfies $\mathrm{E}[(\mu - X)^+] = \mathrm{E}[(X - \mu)^+]$ by its definition. Then we can deduce that $g < \mu$ implies $\mu < h$. We can therefore re-write the optimality curve of $\mathcal{M}$ as

$$f_{\mathcal{M}}(y) = \begin{cases} y & \text{if } y < g \\ \mathrm{E}[\min(y, \max(g, X))] & \text{if } y \in [g, h] \\ \mu & \text{if } y > h \end{cases}$$

From this, we obtain the following characterization of the surrogate costs. We note that the same result is proven by [Scully and Doval, 2024], as the optimality curve for the alternative $\mathcal{M}$ coincides with the optimality curve of a Pandora's Box with optional inspection in the minimization setting.

FACT 8.5. *The water filling surrogate cost $W^*_{\mathcal{M}}$ of $\mathcal{M}$ corresponds to sampling $x \sim X$ and returning*

$$\rho^*(x) := \min(h, \max(g, x)).$$

Committing policies. Now, consider any commitment $\pi \in \mathcal{C}(\mathcal{M})$. Observe that π corresponds to a protocol that determines in advance a decision tree (or a distribution over decision trees) over pre-specified thresholds against which it will weigh X, resulting in a Markov chain $\mathcal{M}^\pi$. Importantly, any commitment will end up partitioning the support $\mathcal{X}$ of distribution $\mathcal{D}$ into a set of intervals, corresponding to its terminal states. We use $\mathcal{I}^\pi$ to denote this set of intervals, and $t(I)$ to denote the terminal state of $\mathcal{M}^\pi$ corresponding to interval $I \in \mathcal{I}^\pi$. Furthermore, the probability of running the Markov chain $\mathcal{M}^\pi$ and ending up in a terminal state $t(I)$ will be precisely $\Pr[X \in I]$. This allows us to obtain the following characterization of surrogate costs for committing policies.

FACT 8.6. *For any committing policy $\pi \in \mathcal{C}(\mathcal{M})$, the water filling surrogate cost $W^*_{\mathcal{M}^\pi}$ corresponds to sampling $x \sim X$ and returning*

$$\rho^\pi(x) := W^*_{\mathcal{M}^\pi}(t(I))$$

for the unique interval $I \in \mathcal{I}^\pi$ that contains x.

We are now ready to prove Lemma 8.4. Let π be the committing policy described in Lemma 8.4. We have already handled the case of $g > \mu$. Next, consider $g > M$. In this case π commits to no weighings, $\mathcal{M}^\pi$ is simply a terminal state of value μ and thus $\rho^\pi(x) = \mu$ for all $x \in \mathcal{X}$. Since $\rho^*(x) \geq g$ for all $x \in \mathcal{X}$, this trivially implies a $\alpha = \mu/g \leq \mu/M$ pointwise approximation and the lemma follows.

If $M \geq g$, then π corresponds to the one-sided halving algorithm with $t_1 = g$ and $t_2 = \min(M, h)$. Observe that by definition, the minimum threshold used by the policy will be some $t_f \in [g, 2g]$. Thus, the first interval in $\mathcal{I}^\pi$ will be $I_0 = (-\infty, t_f]$. Note that

$$\Pr[X \in I_0] = \Pr[X \leq t_f] \geq \Pr[X \geq g] \geq \frac{c}{g}$$

where the last inequality follows from the fact that $c = \mathrm{E}[(g - X)^+] \leq g \cdot \Pr[X \leq g]$. Furthermore, recall that by the definition of (any) amortization, the cost shares for the amortization of a state s satisfy

$$p(s) \cdot c(s) = \sum_{t \in T(S)} p(t) \cdot b_{st}$$

and since $\Pr[X \in I_0] \geq c/g$ and all the action costs are $c(s) = c$, this implies that in any step during the amortization of $\mathcal{M}^\pi$, the surrogate cost of terminal $t(I_0)$ increases by at most g. Finally, we note that the horizon of $\mathcal{M}^\pi$ will be at most

$$k := \log \frac{t_2}{t_f} \leq \log \frac{M}{g} \leq \log \frac{2\mu}{g}$$

and thus we conclude that the total number of amortization steps will be k, and that the total increase in the value of the first terminal will be at most kg.

Up next, we will argue that the terminal $t(I_0)$ will be the terminal that suffers the maximum increase during the water-filling amortization of $\mathcal{M}^\pi$. Note that by definition of $\mathcal{M}^\pi$, the initial value of a terminal state $t(I)$ corresponding to some $I \in \mathcal{I}^\pi$ will be precisely $\mu(I) := \mathrm{E}[X|X \in I]$. Thus, terminal $t(I_0)$ starts with the minimum value. Furthermore, by the one-sided structure of $\mathcal{M}^\pi$, observe that $t(I_0)$ is a terminal state for all intermediate action states, and thus it will participate in all the stages of the water filling amortization. These two facts immediately prove the claim.

From the above, we can summarize that for all $I \in \mathcal{I}^\pi$ we have

$$W^*_{\mathcal{M}^\pi}(t(I)) \leq \mu(I) + k \cdot g$$

and thus for any $x \in \mathcal{X}$, we have $\rho^\pi(x) \leq \mu(I) + kg$ for the unique $I \in \mathcal{I}^\pi$ for which $x \in I$. It remains to upper bound the expectations $\mu(I)$. For $I_0 = (-\infty, t_f] \subseteq (-\infty, 2g]$ we have that $\mu(I_0) \leq 2g$. For the maximum interval $I = (t_2, \infty)$, we have that $\mu(I) \leq 2\mu$; this is a consequence of the fact that $t_2 \leq M$ and thus $\mathrm{E}[X|X > t_2] \leq \mathrm{E}[X|X \geq M] \leq 2\mu$. Finally, for any other interval I, we know that the ratio between its endpoints will be precisely 2 due to the halving and thus $\mu(I) \leq 2x$ for any $x \in I$.

To summarize, we have shown that for all $x \in \mathcal{X}$:

$$\rho^\pi(x) \leq \begin{cases} kg + 2g & \text{if } x < g \\ kg + 2x & \text{if } x \in [g, t_2] \\ kg + 2\mu & \text{if } x > t_2 \end{cases}$$

Observe that since $t_2 \leq M \leq 2\mu$, we have that for any $x \in \mathcal{X}$,

$$\rho^\pi(x) \leq u(x) := kg + 2 \cdot \min(2\mu, \max(x, g)).$$

Notice that the upper bound $u(x)$ is non-decreasing. Now, recall that $\rho^*(x) = \min(h, \max(g, x))$ is also a non-decreasing mapping. This, policy π will α-pointwise approximate $\mathcal{M}$ for

$$a = \max_{x \in X} \frac{u(x)}{\rho^*(x)}$$

and since $\mu \leq h$ and $k = O(\log \frac{\mu}{g})$ this implies a $O(\log \frac{\mu}{g} + \frac{\mu}{M})$-pointwise approximation. ∎

8.2 Proving the Lower Bound. Finally, we prove the lower bound of Theorem 8.2. For ease of notation, let $k := \alpha + 1$ and $B := 2^{k^2}$. We consider the alternative with weighing cost $c = 1$ and random cost X that is continuously distributed in interval $[1, B]$ and has two point masses on 0 and $(k+1)B$, namely:

$$X = \begin{cases} 0 & \text{with probability } 1 - \frac{1}{k} \\ x \in [1, B] & \text{with density } f(x) = \frac{1}{kx^2} \\ (k+1)B & \text{with probability } \frac{1}{kB} \end{cases}$$

Note that since $B > k > 1$ and $\int_{x=1}^{B} \frac{1}{kx^2} dx = \frac{1}{k} - \frac{1}{kB}$, X is indeed a valid random variable. We proceed by computing the relevant parameters of X.

- The expected value of X is $\mu = \mathrm{E}[X] = k + 1 + \frac{1}{k}$ by definition. Thus, $\mu \in [k+1, k+2]$.
- The g-index of $\mathcal{M}$ satisfies $g \in [1, \frac{k}{k-1}]$. Define $\mathrm{exc}(z) := \mathrm{E}[(z - X)^+]$. The claim follows by noting that $\mathrm{exc}(\cdot)$ is non-decreasing; $\mathrm{exc}(1) = 1 - \frac{1}{k} < 1$ and $\mathrm{exc}(\frac{k}{k-1}) > 1$; whereas $\mathrm{exc}(g) = 1$ by definition.

- The h-index of $\mathcal{M}$ is $h = B$, since $\mathrm{E}\left[(X-B)^+\right] = \frac{1}{kB} \cdot [(k+1)B - B] = 1$.

By definition of $\rho^*(x)$, we have that $\rho^*(0) = g$, $\rho^*((k+1)B) = B$ and $\rho^*(x) = \max(g, x)$ for all $x \in [1, B]$. We will now show that no committing policy can achieve an α'-pointwise approximation for the alternative $(X, 1)$ for any $\alpha' < \alpha$.

Fix any commitment π and let $\ell(x) := \mathrm{E}\left[X \in I\right]$ where $I \in \mathcal{I}^\pi$ is the unique interval of the policy's partition that contains x. Clearly, $\ell(x)$ is a lower bound to $\rho^\pi(x)$ and it is also a non-decreasing function. Thus, in order to prove impossibility of pointwise approximation for any $\alpha' < \alpha$, it suffices to show that there exists some $x \in \mathcal{X}$ for which

$$\frac{\ell(x)}{\rho^*(x)} \geq \alpha.$$

We begin by considering the committing policy π that does not perform any weighings. Then, we simply have $\ell(x) = \rho^\pi(x) = \mu$ for all $x \in \mathcal{X}$ and since $\rho^*(x) \geq g$ for all $x \in X$, and the claim follows from $\alpha \leq \mu/g$. Next, consider any commitment that performs weighings, and let t be the maximum threshold that it uses. Clearly, $t < (k+1)B$ otherwise there is no point in the weighing. This means that the final interval in $\mathcal{I}^\pi$ will be $I_\infty := (t, \infty)$. Let $\mu_\infty := \mu(I_\infty) = \mathrm{E}\left[X | X > t\right]$. We distinguish between the following cases:

- If $t \geq B$, then $\mu_\infty = (k+1)B$ and $\rho^*(t) \leq h = B$. In that case, the ratio is at least $k > \alpha$.
- If $t \leq 1$, then $\mu_\infty \geq \mathrm{E}\left[X | X \geq 1\right] = k^2 + k + 1$ and $\rho^*(t) \leq \rho^*(1) = \min(h, \max(g, 1)) = g$. In that case, the ratio is at least $k^2 - k > \alpha$.
- Finally, if $t \in (1, B)$, then $\Pr\left[X \geq t\right] = \frac{1}{kt}$ and thus $\mu_\infty \geq kt$. Since $\rho^*(t) = \max(g, t)$, the ratio is at least $k - 1 = \alpha$.

Thus, in any case there exists some x for which $\ell(x) \geq \rho^*(x)$. As already mentioned, by the fact that both mappings are non-decreasing and by definition of pointwise-approximation, this proves Theorem 8.2.

9 Pandora's Box with Optional Inspection (Maximization)

Finally, we consider the maximization version of *Pandora's Box with Optional Inspection* (henceforth, PBOI), arguably the most studied Pandora's Box variant. In PBOI, the decision-maker may select a box without opening it; we refer to this action as "grabbing" the box. There is a rich line of research addressing PBOI; we provide a detailed discussion of these results in Section 2. In the single-selection setting, PBOI is NP-hard but admits a PTAS [Fu et al., 2022, Beyhaghi and Cai, 2022].

For matroid constraints, [Beyhaghi and Kleinberg, 2019] show that the commitment gap is at least 0.63. While the arguments presented in their work do not suffice for an efficient approximation matching this bound, the same ideas can be extended to match this guarantee; see our discussion in the Introduction for more details. Importantly, their framework requires solving an instance of stochastic submodular maximization, a problem first studied by [Asadpour and Nazerzadeh, 2016]. While this problem can be solved efficiently, the solution is obtained by iteratively solving and rounding an appropriate relaxation. Furthermore, the approximation factor of 0.63 is a direct consequence of their approach, and it is not clear whether instance-dependent guarantees can be obtained. On the other hand, the committing policies obtained through local approximation are significantly more simple and essentially correspond to flipping a random coin for each of the boxes. Furthermore, the achieved approximations naturally depend on the parameters of the boxes, leading to instance-dependent guarantees that are potentially bounded away from the worst-case instances.

As Beyhaghi and Kleinberg [2019] observe, the optimal policy that does not grab any box will attain half of the utility from boxes that the optimal doesn't grab, and the optimal policy that only grabs boxes will attain half of the utility from boxes that the optimal grabs. Thus, uniformly randomizing between these policies will immediately give us a 0.5-approximation.[10] In this section, we use an extension of local approximation to obtain a different randomization between opening and grabing the boxes, obtaining the following result.

[10] Beyhaghi and Kleinberg [2019] make this observation in the single-item setting, but combining it with results of Singla [2017] generalizes it to the matroid setting.

THEOREM 9.1. *There exists an efficient randomized committing policy that achieves a* 0.582*-approximation to the optimal adaptive policy for any instance of matroid-max-PBOI.*

In Section 9.1 we cast PBOI into our CICS framework by considering each optional-inspection box as an MDP. A reasonable first approach would be to establish a local approximation condition for these MDPs in order to prove Theorem 9.1; however, we show that for any $\epsilon > 0$, there exist boxes for which we cannot do better than a $(0.5 + \epsilon)$-local approximation (Section 9.2). We use the intuition from these hard instances to develop a refinement of local approximation, which we call *semilocal approximation*; we then show that, similarly to local approximation, semilocal approximation guarantees can be composed into a (global) approximation ratio (Section 9.3). Finally, in Section 9.4 we show that this new notion of approximation suffices to establish the bound of Theorem 9.1.

Limitations of Semilocal Approximation. There are two prices we pay in refining local approximation to semilocal approximation. First, the definition as we state it currently is specific to PBOI, though it could perhaps be generalized to other MDP families in the future. Second, our composition result for semilocal approximations only holds for matroid feasibility constraints, whereas past works showed results compatible with any frugal (roughly, "greedy") algorithm [Singla, 2017, Gupta et al., 2019, Scully and Doval, 2024]. We are not sure whether this second limitation is fundamental or might be overcome using a different proof technique.

9.1 PBOI Preliminaries. We begin by establishing the necessary notation for PBOI. All the definitions mentioned in this section were previously established for the minimization setting by [Scully and Doval, 2024]; we simply restate them here for the maximization setting.

Fix any optional-inspection PB $\mathcal{B} = (\mathcal{D}, c)$. We use $X \sim \mathcal{D}$ to denote the value of the box and $\mu = \mathrm{E}[X]$ to denote its expectation. We can model $\mathcal{B}$ as a Costly Information MDP $\mathcal{M}$ with two actions: opening the box at a cost of c and transitioning to a terminal state of reward $X \sim \mathcal{D}$, and grabbing the box at a cost of 0 and transitioning to a terminal state of reward μ (Figure 8). Therefore, the set of (deterministic) local commitments for $\mathcal{M}$ consists of just two options $\{o, g\}$ and the optimality curve of $\mathcal{M}$ can be written as $f_{\mathcal{M}}(y) := \max\{f_{\mathcal{M}^o}(y), f_{\mathcal{M}^g}(y)\}$, where

$$f_{\mathcal{M}^o}(y) := \max\{y, \mathrm{E}[\max(X, y)] - c\} \quad , \quad f_{\mathcal{M}^g}(y) := \max\{y, \mu\}$$

are (by definition) the optimality curves under the two commitments. Using the inherent connection between optimality curves and surrogate costs, we can use these expressions to recover the following [Doval, 2018]:

DEFINITION 9.2. (SURROGATE COSTS FOR PBOI) *Let $\mathcal{B} = (\mathcal{D}, c)$ be any optional-inspection box. We define its Gittins index g as the solution to equation $c = \mathrm{E}_{X\sim\mathcal{D}}[(X - g)^+]$ and its backup index as $h = \max(\mu, h')$, where h' is the solution to equation $c = \mathrm{E}_{X\sim\mathcal{D}}[(h' - X)^+]$. Then:*

1. $W_o^* := W_{\mathcal{M}^o}^* = \min\{X, g\}$ *satisfies* $f_{\mathcal{M}}^o(y) = \mathrm{E}[\max\{W_o^*, y\}]$.
2. $W_g^* := W_{\mathcal{M}^g}^* = \mu$ *satisfies* $f_{\mathcal{M}}^g(y) = \mathrm{E}\left[\max\{W_g^*, y\}\right]$.
3. $W^* := W_{\mathcal{M}}^* = \max\{W_o^*, h\}$ *satisfies* $f_{\mathcal{M}}(y) = \mathrm{E}[\max\{W^*, y\}]$.

These indices have an intuitive interpretation: consider the local game $(\mathcal{M}, y)$ for some value of the outside option y. For small values of y, the optimal policy will immediately grab the box, and for large values of y it will accept the outside option. The Gittins index g corresponds to the maximum value of the outside option for which opening is preferable to stopping, and h' corresponds to the minimum value of the outside option for which opening is preferable to grabbing. In other words, the optimal policy for the local game $(\mathcal{M}, y)$ will grab the box if $y \in [0, h']$, it will open the box if $y \in [h', g]$ and it will accept the outside option if $y \geq g$. Note that if $g < h'$, opening the box is never the optimal action. We refer to such boxes $\mathcal{B} = (\mathcal{D}, c)$ as degenerate, and since they are never opened by any optimal policy, we can substitute them by normalized boxes $\mathcal{B} = (\mu, 0)$ without loss of any generality. From now on, we will be assuming that all degenerate boxes have been normalized; this in turn implies that $h = h' \leq \mu$ and $c \leq \mu \leq g$ (Figure 8).

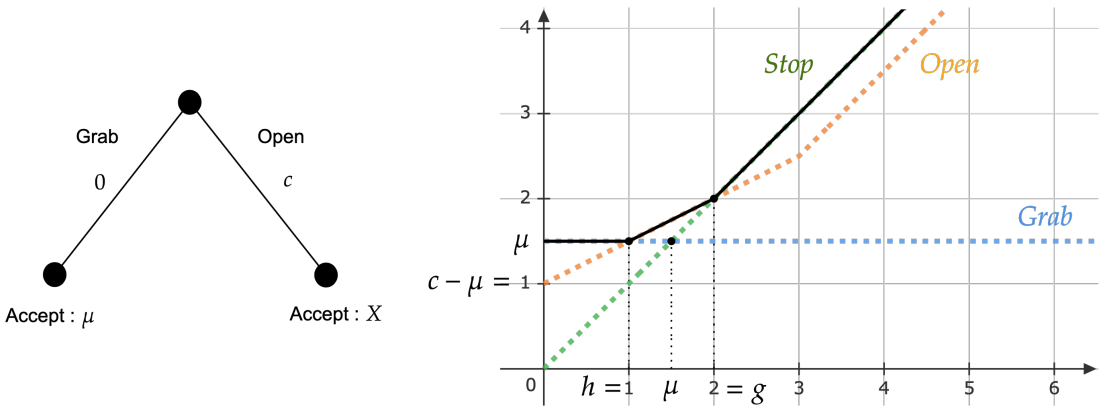


Figure 8: An optional inspection box $\mathcal{B} = (\mathcal{D}, c)$ can me modeled as an MDP $\mathcal{M}$ with two actions, grab and open, incuring costs 0 and c and resulting to rewards $\mathrm{E}[X]$ and X respectively. On the right side, we demonstrate the optimality curves for the local game $(\mathcal{M}, y)$ for a box with cost $c = 0.5$ and reward $X = 3 \cdot \mathrm{Be}(1/2)$.

9.2 Insufficiency of Local Approximation. Since any MDP $\mathcal{M}$ describing an optional-inspection box only has two actions, any (randomized) commitment $\pi \in \mathcal{C}(\mathcal{M})$ is fully described by the probability $p \in [0,1]$ of committing to the grab action. Therefore, by Definition 9.2, the α-local approximation condition for max-PBOI translates to

$$\exists p \in [0,1]: \quad (1-p) \cdot \mathrm{E}[\max\{W_o^*, y\}] + p \cdot \mathrm{E}\left[\max\{W_g^*, y\}\right] \geq \mathrm{E}[\max\{\alpha W^*, y\}].$$

We will now demonstrate a specific box where no commitment $p \in [0,1]$ can achieve a local approximation factor that is better than $1/2$.

Example 9.3. *For any $n \geq 1$, consider the optional-inspection box $\mathcal{B}^n = (\mathcal{D}, c)$ where $c = n - 1$ and*

$$\mathcal{D} = \begin{cases} 1 & w.p.\ 1 - \frac{1}{n^2} \\ n^3 & w.p.\ \frac{1}{n^2} \end{cases}.$$

Straightforward computations show that the parameters of this box are $\mu = n + 1 - n^{-2}$, $g = n^2$ and $h = 1 + n^2/(n+1)$. Using these expressions and the definition of surrogate costs (Definition 9.2), we can proceed to write-down the local approximation condition at points $y = 0$ and $y = \mu$ as follows:

$$\textit{Local approximation at } y = 0\text{:} \quad \alpha \leq 1 - (1-p) \cdot \frac{n^3 - n^2}{n^3 + n^2 - 1}$$

$$\textit{Local approximation at } y = \mu\text{:} \quad \alpha \leq 1 - p \cdot \frac{n^2 - n - 1 + n^{-2}}{n^2}$$

Just from these two values of y, we obtain that as $n \to \infty$, we have that $\alpha \leq 1 - \max(p, 1-p) \leq 0.5$.

To better understand this "worst-case" behavior that Example 9.3 demonstrates, it will be illustrative to visualize the optimality curves of the box $\mathcal{B}^n$ for some large value of n. Figure 9 shows the curves for $n = 20$. We see that when y is small, we have a strong preference for grabbing the box, and when y is larger, we have a very slight preference for opening the box. Intuitively, we might guess that committing to the grabbing action performs quite well, since in the worst case there is only a small *additive* sub-optimality factor. However, local approximation requires a small *multiplicative* sub-optimality factor, which, as Example 9.3 above demonstrates, we cannot achieve.

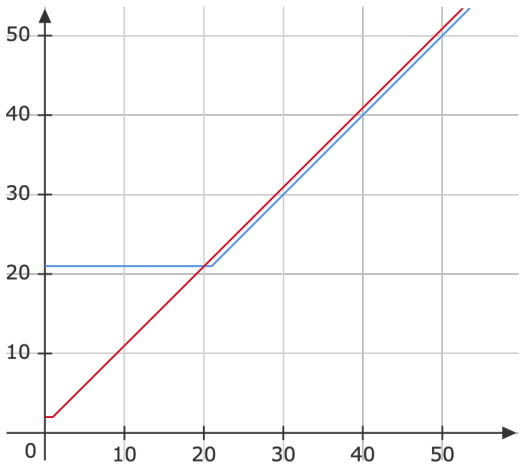


Figure 9: The red and blue lines represent the optimality curves for grabbing $\mathcal{B}^{20}$ and opening $\mathcal{B}^{20}$, respectively. This best possible local approximation achieved for this box is 0.537.

9.3 Semilocal Approximation. The above discussion suggests that we need to reason about additive suboptimality, not just multiplicative. We do so by introducing (α, β)-*semilocal approximation*, defined below. Roughly speaking, the α captures multiplicative suboptimality, while β captures the additive suboptimality. Formally:

Definition 9.4. (Semilocal Approximation) *We say that an optional-inspection box $\mathcal{B} = (\mathcal{D}, c)$ admits an* (α, β)-semilocal approximation *if there exists a probability $p \in [0,1]$ such that for all $y \in \mathbb{R}$:*

$$(1-p) \cdot \mathrm{E}\left[\max(W_o^*, y)\right] + p \cdot \mathrm{E}\left[\max(W_g^*, y)\right] \geq \mathrm{E}\left[\max(\alpha W^*, y)\right] - p\beta\mu.$$

Re-writing the above expression as

$$(1-p) \cdot \mathrm{E}\left[\max(W_o^*, y)\right] + p \cdot \left(\mathrm{E}\left[\max(W_g^*, y)\right] + \beta\mu\right) \geq \mathrm{E}\left[\max(\alpha W^*, y)\right],$$

it becomes clear that (α, β)-semilocal approximation is basically an α-local approximation where we have boosted the value attained from grabbing by an additive factor proportional to the mean. It is important that we do this only for grabbing—boosting the value of inspecting by an additive factor would lead to poor guarantees when composing semilocal approximations in the matroid selection setting.

Equipped with the semilocal approximation definition, we now show that semilocal approximations can be composed under matroid feasibility constraints:

Theorem 9.5. *Let $\alpha > \beta \geq 0$, and let $\mathbb{I} = (\mathcal{B}_1, \ldots, \mathcal{B}_n, \mathcal{F})$ be any instance of matroid-max-PBOI where each constituent box $\mathcal{B}_i$ admits an (α, β)-semilocal approximation under some probability $p_i \in [0,1]$. Then,* $\mathrm{CG}(\mathbb{I}) \geq \alpha - \beta$.

Algorithm 9.6. (Semilocal Approximation Composition Algorithm)

Input. A normalized matroid-max-PBOI instance $\mathbb{I} = (\mathcal{B}_1, \ldots, \mathcal{B}_n, \mathcal{F})$ and a vector of probabilities $(p_1, \ldots, p_n)$.
$S^{\mathrm{grab}} \leftarrow \{\}$. ▷ set of boxes marked as "grab"
Relabel the boxes such that $\mu_1 \geq \cdots \geq \mu_n$.
for $i = 1, 2, \ldots, n$ **do**
    $L_i \leftarrow 1$ **if** $S^{\mathrm{grab}} \cup \{i\} \in \mathcal{F}$ **else** 0 ▷ $L_i = 0$ means we never want to grab box i
    Sample $K_i \leftarrow \mathrm{Bernoulli}(p_i)$ ▷ $K_i = 1$ means *provisionally* mark box i "grab"
    If $K_i = 1$ **and** $L_i = 1$ **then** $S^{\mathrm{grab}} \leftarrow S^{\mathrm{grab}} \cup \{i\}$. ▷ fully mark box i as "grab"
Commit to grabbing the boxes S^{grab} and opening the boxes in $[n] \setminus S^{\mathrm{grab}}$.
Run the optimal (index) policy under the resulting commitment.

The committing policy achieving the bound of Theorem 9.5 is presented in Algorithm 9.6. Importantly, this policy does not commit to grabbing each box $\mathcal{B}_i$ with probability p_i independently (as would be the case when composing local approximation guarantees). Instead, it examines the boxes in decreasing mean order, flips a coin with probability p_i for each box $\mathcal{B}_i$ and only commits to grabbing it if (i) the coin flip succeeds and (ii) the set of boxes that are committed to the grab action after inserting $\mathcal{B}_i$ remains an independent set of the underlying matroid. Once the commitments for each box have been specified, we simply run the optimal policy for the resulting matroid-max-CICS instance over Markov chains (equivalently, we substitute each grab box $(\mathcal{D}, c)$ with a box $(\mu, 0)$ and run Weitzman's algorithm for the mandatory inspection setting). Therefore, as long as the probabilities p_i achieving the semilocal approximation can be efficiently computed, the entire algorithm runs in polynomial time. To ease the presentation, we defer the proof of Theorem 9.5 to Section 9.5

We note that both the definition and composition algorithm for semilocal approximation are specifically tailored to the PBOI model, though we believe modest extensions beyond PBOI may be possible. The main limitaiton is that the proof of Theorem 9.5 relies crucially on the fact that if we choose to grab a box, then the box's terminal value is deterministic, namely μ. It is thus natural to try extending semilocal approximation to other Pandora's box variants that have a grab-like action that yields a deterministic terminal value, but we leave this to future work.

9.4 Breaking the 0.5 Barrier. We are finally ready to prove Theorem 9.1. From Theorem 9.5, it suffices to prove that all optional inspection boxes admit an (α, β)-semilocal approximation for some α, β such that $\alpha - \beta \geq 0.582$. This is achieved by the following lemma.

LEMMA 9.7. *Let $\mathcal{B}$ be any (non-degenerate) box with opening cost c and mean value μ. For any $\beta \geq 0$, let*

$$\alpha(\beta) := \begin{cases} \frac{1}{1+\frac{c}{\mu}-\beta\frac{c}{\mu-c}} & \text{if } \frac{1}{1+\frac{c}{\mu}-\beta\frac{c}{\mu-c}} \in (0,1]. \\ 1 & \text{otherwise.} \end{cases}$$

Then, the probability $p = \frac{c}{\mu}\alpha(\beta)$ achieves an $(\alpha(\beta), \beta)$-semilocal approximation for $\mathcal{B}$.

Proof. To simplify notation we denote $\alpha(\beta)$ as α throughout the proof. By Definition 9.4, proving this lemma amounts to showing that these α and β satisfy:

$$(1-p)\,\mathrm{E}\left[\max\left(W_o^*, y\right)\right] + p\max\left(\mu, y\right) \geq \mathrm{E}\left[\max\left(\alpha W^*, y\right)\right] - p\beta\mu \tag{9.10}$$

for all $y \geq 0$. This is equivalent to showing that

$$f(y) := (1-p)\,\mathrm{E}\left[\max\left(W_o^*, y\right)\right] + p\max\left(\mu, y\right) - \mathrm{E}\left[\max\left(\alpha W^*, y\right)\right] + p\beta\mu$$

satisfies $f(y) \geq 0$ for all $y \geq 0$. We will first show that it is sufficient to check that $f(0) \geq 0$ and that $f(y) \geq 0$ for $y \geq \mu$. It follows from Definition 9.2 that at all values y for which $f'(y)$ is defined, it is equal to

$$(1-p)\Pr\left[y \geq W_o^*\right] + p\mathbb{1}(y > \mu) - \mathbb{1}(y > \alpha h)\Pr\left[y > \alpha W_o^*\right].$$

Observe that

- if $y < \alpha h < \mu$, then $f'(y) \geq 0$,
- if $\alpha h < y < \mu$, then since $\alpha \leq 1$, we must have $f'(y) \leq 0$.

Thus a global minimum of $f(y)$ must be at $y = 0$ or on $y \geq \mu$, so it is sufficient to check that Equation (9.10) is satisfied when $y = 0$ and $y \geq \mu$. When $y = 0$, Equation (9.10) reduces to

$$(1-p)(\mu - c) + p\mu \geq \alpha\mu - p\beta\mu.$$

which holds for the values of α and p given in the lemma. When $y \geq \mu$, Equation (9.10) reduces to

$$(1-p)\,\mathrm{E}\left[\max\left(W_o^*, y\right)\right] + py \geq \mathrm{E}\left[\max\left(\alpha W^*, y\right)\right] - p\beta\mu = \alpha\,\mathrm{E}\left[\max\left(W_o^*, \frac{y}{\alpha}\right)\right] - p\beta\mu \tag{9.11}$$

where the last equality follows from the fact that $\mathrm{E}\left[\max\left(\alpha W^*, y\right)\right] = \max\{\mathrm{E}\left[\max\left(\alpha W_o^*, y\right)\right], \alpha\mu\}$. The slope of $\mathrm{E}\left[\max\{W_o^*, y\}\right]$ is bounded above by 1, so

$$\alpha\,\mathrm{E}\left[\max\left(W_o^*, \frac{y}{\alpha}\right)\right] \le \alpha\,\mathrm{E}\left[\max\left(W_o^*, y\right)\right] + \alpha\left(\frac{y}{\alpha} - y\right).$$

Applying this bound to Equation (9.11), we find that Equation (9.11) holds if,

$$(\alpha + p - 1)(\mathrm{E}\left[\max\left(W_o^*, y\right)\right] - y) \le p\beta\mu. \tag{9.12}$$

Since the slope of $\mathrm{E}\left[\max\{W_o^*, y\}\right]$ is bounded above by 1, it must be that $\mathrm{E}\left[\max\left(W_o^*, y\right)\right] - y$ is maximized at $y = 0$ where it is equal to $\mu - c$, so Equation (9.12) holds if,

$$(\alpha + p - 1)(\mu - c) \le p\beta\mu.$$

Observe that if $p = \alpha\frac{c}{\mu}$, we can rewrite this inequality as

$$\alpha\left(1 + \frac{c}{\mu} - \frac{c}{\mu - c}\beta\right) \le 1,$$

which is satisfied by the α given in the lemma. ∎

By optimizing over $\alpha(\beta) - \beta$, we instantiate Lemma 9.7 with $\beta = 0.1$, to obtain the corresponding probability. It is straightforward to verify (using a computer algebra system or numerical solver) that since $\frac{c}{\mu} \in [0, 1]$ (recall that we have normalized all degenerate boxes), we have

$$\left(1 + \frac{c}{\mu} - \frac{c}{10(\mu - c)}\right)^{-1} \ge 0.682$$

whenever $(1 + \frac{c}{\mu} - \frac{c}{10(\mu-c)})^{-1} \in (0, 1]$. This means $\alpha(0.1) \ge 0.682$, and thus all boxes admit a $(0.682, 0.1)$-semilocal approximation, from which the proof of Theorem 9.1 follows. Critical to the above argument is that Lemma 9.7 implies there exists a *universal* pair (α, β), namely $(0.682, 0.1)$, such that all boxes admit an (α, β)-semilocal approximation. It would *not* suffice to show only that for each box, there exists a *box-specific* pair (α', β') such that the box admits a (α', β')-semilocal approximation, even if we always had $\alpha' - \beta' \ge 0.582$. This is because Theorem 9.5 requires all boxes to admit a (α, β)-approximation for the same pair (α, β).

9.5 Composition of Semi-Local Approximation. We will finally analyze the composition Algorithm 9.6 and prove Theorem 9.5. To simplify the notation, we let $W_i^{\mathrm{open}} := W_{o,i}^*$ be the surrogate cost of the opening action for box $\mathcal{B}_i$. There are two main steps of the proof, each stated and proved in a lemma below. We express both steps in terms of the random variables

$$W_i^{\mathrm{alg}} = \begin{cases} W_i^{\mathrm{open}} & \text{if } K_i L_i = 0 \\ \mu_i & \text{if } K_i L_i = 1, \end{cases}$$

where K_i and L_i are as defined in Algorithm 9.6. One can think of W_i^{alg} as the surrogate value of box i conditional on the state of the algorithm after line 7. The first step, Lemma 9.8, is to express the value achieved by Algorithm 9.6 in terms of W_i^{alg}:

LEMMA 9.8. *For any max-matroid Pandora's box instance $\mathbb{I} = (\mathcal{B}_1, \ldots, \mathcal{B}_n, \mathcal{F})$ and any vector of probabilities $(p_1, \ldots, p_n)$, the expected value achieved by Algorithm 9.6 is*

$$\mathrm{E}\left[\text{value achieved by Algorithm 9.6}\right] = \mathrm{E}\left[\max_{S \in \mathcal{F}} \sum_{i \in S} W_i^{\mathrm{alg}}\right].$$

The second step, Lemma 9.9, is to compare the resulting expression to an upper bound on the optimal value. This step uses the semilocal (α, β)-approximation guarantee from Definition 9.4, which gives us a relationship between W_i^{alg} and W_i^*.

LEMMA 9.9. *Under the hypotheses of Theorem 9.5,*

$$\mathrm{E}\left[\max_{S\in\mathcal{F}}\sum_{i\in S}W_i^{\mathrm{alg}}\right]\geq\mathrm{E}\left[\max_{S\in\mathcal{F}}\sum_{i\in S}\alpha W_i^*-\sum_{i\in S^{\mathrm{grab}}}\beta\mu_i\right],$$

where S^{grab} refers to the value of the S^{grab} variable in Algorithm 9.6.

Combining the two lemmas yields

$$\mathrm{E}\left[\text{value achieved by Algorithm 9.6}\right]\geq\mathrm{E}\left[\max_{S\in\mathcal{F}}\sum_{i\in S}\alpha W_i^*-\sum_{i\in S^{\mathrm{grab}}}\beta\mu_i\right],$$

where S^{grab} is the set of boxes marked "grab" at the after line 7 of the algorithm. It remains only to relate the two terms on the right-hand side to the optimal expected value.

- From the analogue of Theorem 4.9 for the maximization setting (see Theorem A.7 in Section A), we know that the water filling surrogate costs provide an upper bound on the utility of the optimal adaptive policy for any max-CICS instance. Thus:

$$\mathrm{E}\left[\max_{S\in\mathcal{F}}\sum_{i\in S}W_i^*\right]\geq\mathrm{E}\left[\text{value of optimal policy}\right].$$

- Because Algorithm 9.6 ensures $S^{\mathrm{grab}}\in\mathcal{F}$ by construction, the following policy is feasible: "Compute S^{grab} as in Algorithm 9.6, but then simply grab the boxes in S^{grab}." This algorithm achieves value $\mathrm{E}\left[\sum_{i\in S^{\mathrm{grab}}}\mu_i\right]$, which means

$$\mathrm{E}\left[\sum_{i\in S^{\mathrm{grab}}}\mu_i\right]\leq\mathrm{E}\left[\text{value of optimal policy}\right].$$

Therefore, as desired, $\mathrm{E}\left[\text{value achieved by Algorithm 9.6}\right]\geq(\alpha-\beta)\cdot\mathrm{E}\left[\text{value of optimal policy}\right]$ which implies a lower bound of $\alpha-\beta$ for the commitment gap.

Proof of Lemma 9.8. Consider running Algorithm 9.6 through line 7, but not further. All of the randomness thus far comes from the coin flips K_i, and no boxes have been opened yet. This means that conditional on the coin flips K_i, the expected value achieved is that of a *mandatory-inspection* instance $\mathbb{I}'=(\mathcal{B}_1',\ldots,\mathcal{B}_n',\mathcal{F})$ whose ith box $\mathcal{B}_i'$ is defined as follows:

- If $K_iL_i=0$ (marked "open"), $\mathcal{B}_i'=(\mathcal{D}_i,c_i)$, i.e. the box is the same as the original instance.
- If $K_iL_i=1$ (marked "grab"), $\mathcal{B}_i'=(\mu_i,0)$, i.e. the box is free to open and always contains value μ_i.

Under instance $\mathbb{I}'$, box i's surrogate value is given by W_i^{alg}. Since this is an instance of max-CICS over Markov chains, from the counterpart of Theorem 4.4 for maximization (see Theorem A.4 in Section A) we have

$$\mathrm{E}\left[\text{value achieved by Algorithm 9.6}|K_1,\ldots,K_n\right]=\mathrm{E}\left[\max_{S\in\mathcal{F}}\sum_{i\in S}W_i^{\mathrm{alg}}|K_1,\ldots,K_n\right].$$

The lemma then follows by the law of total expectation. ∎

Proof of Lemma 9.9. We want to prove that

$$\mathrm{E}\left[\max_{S\in\mathcal{F}}\sum_{i\in S}W_i^{\mathrm{alg}}\right]\geq\mathrm{E}\left[\max_{S\in\mathcal{F}}\sum_{i\in S}\alpha W_i^*-\sum_{i\in S^{\mathrm{grab}}}\beta\mu_i\right]. \tag{9.13}$$

The outline of the proof is as follows. We begin with the left-hand side of Equation (9.13). Then, for each box i, we swap W_i^{alg} with αW_i^*, and also subtract $\beta\mu_i$ if box i is marked "grab". Because each box admits a semilocal (α, β)-approximation, each of these replacements only decreases the expression's expected value. After all n replacements, we are left with the right-hand side of Equation (9.13), as desired. This is the same strategy used by Scully and Doval [2024, Theorem 5.4], but some careful conditioning is needed to account for the $\beta\mu_i$ subtractions.

In order to notate the one-by-one replacement outlined above, let

$$W_i^{(j)} = \begin{cases} W_i^{\text{alg}} & \text{if } i \le j \\ \alpha W_i^* & \text{if } i > j, \end{cases} \qquad\qquad U^{(j)} = \max_{S \in \mathcal{F}} \sum_{i \in S} W_i^{(j)}.$$

Using this notation and recalling how S^{grab} is defined in Algorithm 9.6, we can rewrite our goal Equation (9.13) as

$$\mathrm{E}\left[U^{(n)}\right] \ge \mathrm{E}\left[U^{(0)} - \sum_{i=1}^{n} \beta K_i L_i \mu_i\right].$$

Therefore, it suffices to show that for all $j \in \{1, \dots, m\}$,

$$\mathrm{E}\left[U^{(j)}\right] \ge \mathrm{E}\left[U^{(j-1)} - \beta K_j L_j \mu_j\right]. \tag{9.14}$$

We will show Equation (9.14) using the definition of semilocal approximation (Definition 9.4). But in order to do so, we need to express each side in terms of a maximum between W_j^{alg} or αW_j^* and a quantity that is independent of box j's value X_j and coin flip K_j (this will take the place of the outside option in the definition of semilocal approximation). We express the latter quantity in terms of

$$Y_{\neq j} = \max_{S \in \mathcal{F} : j \notin S} \sum_{i \in S} W_i^{(j)}, \qquad\qquad Z_{\neq j} = \max_{S \in \mathcal{F} : j \in S} \sum_{i \in S \setminus \{j\}} W_i^{(j)}.$$

These can both be seen as optimal total surrogate values achievable without box j. The difference is that $Y_{\neq j}$ optimizes over sets that exclude j, whereas $Z_{\neq j}$ optimizes over sets that include j (but still excludes box j's surrogate value from the sum). With the definitions of $Y_{\neq j}$ and $Z_{\neq j}$ in hand, we can express $U^{(j)}$ and $U^{(j-1)}$ as

$$\begin{aligned} U^{(j)} &= \max\left(Y_{\neq j}, Z_{\neq j} + W_j^{\text{alg}}\right) = Z_{\neq j} + \max\left(W_j^{\text{alg}}, Y_{\neq j} - Z_{\neq j}\right), \\ U^{(j-1)} &= \max\left(Y_{\neq j}, Z_{\neq j} + \alpha W_j^*\right) = Z_{\neq j} + \max\left(\alpha W_j^*, Y_{\neq j} - Z_{\neq j}\right). \end{aligned}$$

So to show Equation (9.14), it suffices to show

$$\mathrm{E}\left[\max\left(W_j^{\text{alg}}, Y_{\neq j} - Z_{\neq j}\right)\right] \ge \mathrm{E}\left[\max\left(\alpha W_j^*, Y_{\neq j} - Z_{\neq j}\right) - \beta K_j L_j \mu_j\right].$$

Letting $K_{<j} = (K_1, \dots, K_{j-1})$, by the law of total expectation, it suffices to show

$$\mathrm{E}\left[\max\left(W_j^{\text{alg}}, Y_{\neq j} - Z_{\neq j}\right) | K_{<j}, Y_{\neq j}, Z_{\neq j}\right] \ge \mathrm{E}\left[\max\left(\alpha W_j^*, Y_{\neq j} - Z_{\neq j}\right) - \beta K_j L_j \mu_j | K_{<j}, Y_{\neq j}, Z_{\neq j}\right]. \tag{9.15}$$

The key to showing Equation (9.15) is observing the following independence facts:

- (K_j, X_j) is independent of $K_{<j}$. This is because the coin flips $K_{<j}$ affect neither the coin flip K_j nor the box value X_j.
- (K_j, X_j) is conditionally independent of $(Y_{\neq j}, Z_{\neq j})$ given $K_{<j}$. This is because once $K_{<j}$ are fixed, the values $(Y_{\neq j}, Z_{\neq j})$ are a function of the values of boxes other than j, and the box values are mutually independent.

The main obstacle to applying the semilocal approximation condition to Equation (9.15) is that W_j^{alg} depends on L_j, which in turn depends on $K_{<j}$. Fortunately, we see from Algorithm 9.6 that

$$L_{\le j} = (L_1, \dots, L_j) \text{ is a deterministic function of } K_{<j} = (K_1, \dots, K_{j-1}).$$

This is because for all i, when executing line 7, the only randomness the algorithm has used is the past coin flips $K_{<i}$. So to show Equation (9.15), we split into cases based on whether $L_j = 0$ or $L_j = 1$.

Suppose that $L_j = 1$. More precisely, suppose $K_{<j} = k_{<j}$, where $k_{<j}$ is any bit vector such that $K_{<j} = k_{<j}$ induces $L_j = 1$ in Algorithm 9.6. In this case, the coin flip K_j impacts whether we mark box j as "grab" or "open", so we will use the semilocal approximation guarantee from Definition 9.4. By the assumption on $k_{<j}$ and the fact that $K_j \sim \text{Bernoulli}(p_j)$ independently of $K_{<j}$,

$$\mathrm{E}\left[K_j L_j | K_{<j} = k_{<j}\right] = \mathrm{E}\left[K_j\right] = p_j.$$

So, by Definition 9.4,

$$\begin{aligned}
\mathrm{E}\left[\max\left(W_j^{\text{alg}}, y\right) | K_{<j} = k_{<j}\right] &= (1 - p_j)\,\mathrm{E}\left[\max\left(W_j^{\text{open}}, y\right]\right) + p_j \max\left(\mu_j, y\right) \\
&\geq \mathrm{E}\left[\max\left(\alpha W_j^*, y\right)\right] - \beta p_j \mu_j \\
&= \mathrm{E}\left[\max\left(\alpha W_j^*, y\right) - \beta K_j \mu_j\right] \\
&= \mathrm{E}\left[\max\left(\alpha W_j^*, y\right) - \beta K_j L_j \mu_j | K_{<j} = k_{<j}\right].
\end{aligned}$$

The fact that $(Y_{\neq j}, Z_{\neq j})$ is conditionally independent of (K_j, X_j) given $K_{<j}$ completes the proof of Equation (9.15) on the event $L_j = 1$.

Suppose now that $L_j = 0$. In this case, we mark box j as "open" regardless of the coin flip K_j, so instead of using the semilocal approximation condition, we will show that marking box j as "open" does not lose any any potential value. Specifically, we will show the following:

(a) For all $y \geq h_j$, we have $\mathrm{E}\left[\max\left(W_j^{\text{open}}, y\right)\right] = \mathrm{E}\left[\max\left(W_j^*, y\right)\right]$.

(b) If $L_j = 0$, then $Y_{\neq j} - Z_{\neq j} \geq \mu_j$.

Together with the fact that $\mu_j \geq h_j$ (which holds by the assumption that all degenerate boxes have been normalized), facts (a) and (b) imply that for any $k_{<j}$ such that $K_{<j} = k_{<j}$ induces $L_j = 0$ in Algorithm 9.6,

$$\begin{aligned}
\mathrm{E}\left[\max\left(W_j^{\text{alg}}, Y_{\neq j} - Z_{\neq j}\right) | K_{<j} = k_{<j}\right] &= \mathrm{E}\left[\max\left(W_j^{\text{open}}, Y_{\neq j} - Z_{\neq j}\right) | K_{<j} = k_{<j}\right] \\
&= \mathrm{E}\left[\max\left(W_j^*, Y_{\neq j} - Z_{\neq j}\right) | K_{<j} = k_{<j}\right] \\
&\geq \mathrm{E}\left[\max\left(\alpha W_j^*, Y_{\neq j} - Z_{\neq j}\right) | K_{<j} = k_{<j}\right] \\
&= \mathrm{E}\left[\max\left(\alpha W_j^*, Y_{\neq j} - Z_{\neq j}\right) - \beta K_j L_j \mu_j | K_{<j} = k_{<j}\right],
\end{aligned}$$

which completes the proof of Equation (9.15) on the event $L_j = 0$. It remains only to show (a) and (b). Fact (a) holds by definition: recall that h_j denotes the smallest value of the outside option y for which the optimal policy in the local game will prefer opening to grabbing. For (b), we will show that if $L_j = 0$, then $Y_{\neq j} \geq \mu_j + Z_{\neq j}$. Let

- $B^Z \in \text{argmax}_{S \in \mathcal{F}: j \in S} \sum_{i \in S \setminus \{j\}} W_i^{(j)}$ be a maximizing basis in the definition of $Z_{\neq j}$,
- $S_{<j}^{\text{grab}} = S^{\text{grab}} \cap \{1, \ldots, j-1\}$ be the boxes marked "grab" before the j-th step in the loop, and
- B^{grab} be $S_{<j}^{\text{grab}}$ extended to a basis by elements of B^Z, so that $S_{<j}^{\text{grab}} \subseteq B^{\text{grab}} \subseteq S_{<j}^{\text{grab}} \cup B^Z$.

Because $L_j = 0$, we have $S_{<j}^{\text{grab}} \cup \{j\} \notin \mathcal{F}$, which means $j \notin B^{\text{grab}}$. But $j \in B^Z$ by definition, so $j \in B^Z \setminus B^{\text{grab}}$. By the basis exchange property, there exists $k \in B^{\text{grab}} \setminus B^Z$ such that the following is a basis:

$$B^Y = (B^Z \setminus \{j\}) \cup \{k\}.$$

But $B^{\text{grab}} \setminus B^Z \subseteq S_{<j}^{\text{grab}}$, which means $W_k^{(j)} = \mu_k \geq \mu_j$ since Algorithm 9.6 iterates over the boxes in decreasing order of mean value. This means

$$Y_{\neq j} \geq \sum_{i \in B_Y} W_i^{(j)} = \mu_k + \sum_{i \in B_Z \setminus \{j\}} W_i^{(j)} \geq \mu_j + \sum_{i \in B_Z \setminus \{j\}} W_i^{(j)} = \mu_j + Z_{\neq j}.$$

and the proof is completed. ∎

Appendix

A The Maximization Setting

In this section, we describe how our entire framework extends to the maximization setting under matroid feasibility constraints. Our objective will be to restate our claims from Sections 4 and 5, as they were only established with respect to the matroid-min-CICS problem. Since the proofs follow exactly the same steps, rather than re-proving all our results, we simply discuss the differences, where there are any.

A.1 Amortization for Markov Chains. The cost amortization of a Markov chain is defined in the same manner, with the difference that instead of increasing the terminal values of the trajectories to obtain surrogate costs, we now *decrease* them to obtain *surrogate values*. Furthermore, the index of a state now corresponds to the *maximum* surrogate value among all downwards trajectories instead of the minimum. Formally:

Definition A.1. (Cost Amortization for Maximization) *A cost amortization for any Markov chain* $\mathcal{M} = (S, \sigma, A, c, \mathcal{D}, V, T)$ *is a non-negative vector* $b = \{b_{s\tau}\}_{s\in S, \tau\in\Upsilon(s)}$ *satisfying* $\sum_{\tau\in\Upsilon(s)} p(\tau) b_{s\tau} = p(s)c(s)$ *for all states* $s \in S$. *Based on this amortization, we define:*

- *The amortized value of a trajectory* $\tau \in \Upsilon$ *as* $\rho_b(\tau) := v(\tau) - \sum_{s\in\tau} b_{s\tau_s}$.
- *The surrogate value of the Markov chain* $\mathcal{M}$ *as the random variable* $\rho_{\mathcal{M},b}$ *that takes on value* $\rho_b(\tau)$ *for* $\tau \in \Upsilon$ *with probability* $p(\tau)$.
- *The index of a state* $s \in S$ *of the Markov chain* $\mathcal{M}$ *as* $I_{\mathcal{M},b}(s) := \max_{\tau\in\Upsilon: s\in\tau} \rho_b(\tau)$.

Intuitively, we postpone the payment of the action costs until a terminal state is accepted, in which case a smaller (compared to the original value) surrogate value is collected. From this, we follow precisely the same steps as in the proof of Lemma 4.2 to upper bound the the utility of any algorithm for matroid-max-CICS through the surrogate values.

Lemma A.2. (Markov chain upper upper for maximization settings) *Consider any instance* $\mathbb{I} = (\mathbb{M}, \mathcal{F})$ *of matroid-max-CICS over Markov chains and let* b_i *be any cost amortization of* $\mathcal{M}_i$ *with surrogate value* $\rho_i := \rho_{\mathcal{M}_i, b_i}$ *for all* $i \in [n]$. *Then,* $\mathrm{OPT}(\mathbb{I}) \le \mathrm{E}\left[\max_{S\in\mathcal{F}} \sum_{i\in S} \rho_i\right]$.

Up next, we extend our definition of water filling amortization to the maximization setting. Like before, the **water draining cost amortization** is described algorithmically in a bottom-up fashion where states are considered in the chain in reverse topological order, starting from the terminals up towards the root σ. Trajectories $\tau \in \Upsilon(t)$ for terminal states $t \in T$ are singletons and do not carry cost shares. Consider a state s such that the cost shares for all states reachable from it have been determined. The state distributes its total cost $c(s)$ across its downstream trajectories $\tau \in \Upsilon(s)$, starting from those with the **maximum current value**, until the equation $\sum_{\tau\in\Upsilon(s)} p(\tau) b_{s\tau} = p(s)c(s)$ is satisfied. In other words, instead of increasing the cost of the minimum-cost trajectory, we now decrease the value of the maximum-value trajectory. We use $W^*_{\mathcal{M}}$ to denote the water draining surrogate value of a Markov chain $\mathcal{M}$, and $I^*_{\mathcal{M}}(s)$ to denote the water draining index of a state s in $\mathcal{M}$.

Through this amortization, we once again define the corresponding index based policy; naturally, the policy will now select to advance the feasible Markov chain of *maximum* index.

Definition A.3. (Water Draining Index policy) *The water draining index policy for an instance* $\mathbb{I} = (\mathbb{M}, \mathcal{F})$ *of matroid-max-CICS over Markov chains selects at every step the Markov chain* $i^* = \mathrm{argmax}_{i\in\mathcal{F}_S} I^*_i(s_i)$, *where* s_i *is the current state of Markov chain* $\mathcal{M}_i$; S *is the set of terminated (selected) Markov chains; and* $\mathcal{F}_S = \{i : S \cup \{i\} \in \mathcal{F}\}$.

From the same observations as in the proof of Lemma 4.2, it immediately follows that this algorithm achieves two desired properties: it always selects the feasible set of Markov chains with maximum total surrogate value, and whenever it advances a state $s \in S$ that contributes to the surrogate value of a downwards trajectory $\tau \in \Upsilon(s)$ (i.e. $b_{s\tau} > 0$) and nature realizes τ, the algorithm will select it. From this, and the fact that the algorithm that greedily adds the maximum weight feasible element is optimal for the matroid independent set problem, the following counterpart to Theorem 4.4 follows, establishing the optimality of the Water Draining Index policy for matroid-max-CICS.

THEOREM A.4. *For any matroid-max-CICS instance $\mathbb{I} = (\mathbb{M}, \mathcal{F})$ over Markov chains, the expected utility of the water draining index policy is equal to $\mathrm{E}\left[\max_{S\in\mathcal{F}} \sum_{i\in S} W^*_{\mathcal{M}_i}\right]$. The policy is therefore optimal for $\mathbb{I}$.*

A.2 Optimality Curves. The definition of a local game $(\mathcal{M}, y)$ naturally extends to the maximization setting; at any step, the decision maker can either advance the Markov chain $\mathcal{M}$ at a cost (until it reaches a terminal state in which case it may collect its value and terminate), or collect the reward y of the outside option and terminate. Like before, we use $f_{\mathcal{M}}(y)$ to denote the utility of the optimal adaptive policy in this game. From Theorem A.4, we immediately have that

$$f_{\mathcal{M}}(y) = \mathrm{E}\left[\max\{y, W^*_{\mathcal{M}}\}\right].$$

From this expression, we obtain that the CDF of the water draining surrogate value $W^*_{\mathcal{M}}$ can be derived from the optimality curve as $\frac{d}{dy} f_{\mathcal{M}}(y)$. This in turn allows us to define water draining surrogate values for arbitrary MDPs.

DEFINITION A.5. (WATER DRAINING SURROGATE VALUES FOR MDPS) *Let $\mathcal{M}$ be an MDP with optimality curve $f_{\mathcal{M}}$. The water draining surrogate value for $\mathcal{M}$ is the random variable $W^*_{\mathcal{M}}$ generated from the CDF $\frac{d}{dy} f_{\mathcal{M}}(y)$. That is, $W^*_{\mathcal{M}}$ is the random variable satisfying $f_{\mathcal{M}}(y) = \mathrm{E}\left[\max(y, W^*_{\mathcal{M}})\right]$ for all $y \in \mathbb{R}$.*

A.3 Amortization for MDPs. We will now use the definition of the water draining surrogate values to prove the following counterpart of Lemma 4.10 that characterizes the water draining surrogate value of an MDP. Since this is the most technical proof of the framework and there are a few arguments that change in the maximization setting, we provide a proof sketch for the result.

LEMMA A.6. *For any MDP $\mathcal{M} = (S, \sigma, A, c, \mathcal{D}, V, T)$ and any deterministic commitment $\pi \in \mathcal{C}(\mathcal{M})$, generating a Markov chain $\mathcal{M}^\pi$ with states $S_\pi \subseteq S$, realizable trajectories $\Upsilon_\pi \subseteq \Upsilon$ and a distribution p_π over trajectories and reachable states, there exists an amortized cost function $\rho^\pi : \Upsilon_\pi \mapsto \Delta(\mathbb{R})$ mapping trajectories to distributions over costs and a non-negative cost sharing vector $b^\pi = \{b^\pi_{s\tau}\}_{s\in S_\pi, \tau\in\Upsilon_\pi(s)}$, such that the following properties hold:*

1. **Cost Sharing:** *the amortized cost of a trajectory pays for its own acceptance value and the cost shares it sends to upstream states. For all $\tau \in \Upsilon_\pi$, it holds that $\mathrm{E}\left[\rho^\pi(\tau)\right] = v(\tau) - \sum_{s\in\tau} b^\pi_{s\tau_s}$.*

2. **Cost Dominance:** *the cost shares received by any state pay towards its action cost, but do not overpay. For all $s \in S_\pi$, it holds that $\sum_{\tau\in\Upsilon_\pi(s)} p_\pi(\tau) b^\pi_{s\tau} \le p_\pi(s) c(\pi(s))$.*

3. **Action Independence:** *sampling a trajectory $\tau \sim p_\pi$ and then sampling from the amortized cost distribution $\rho^\pi(\tau)$ generates a random surrogate cost for the MDP. This random variable is distributed identically to the water draining surrogate cost $W^*_{\mathcal{M}}$.*

Proof. Notice that the only change with respect to the statement of Lemma 4.10 for the minimization setting is the negative sign in the cost sharing property. Our proof will mirror the proof of Lemma 4.10, and we use the same notation throughout. Once again, the proof proceeds by induction on the horizon of the MDP; the $H = 1$ case remains trivial.

For the amortization of the action costs $c(a_j)$, we will now define g_j to be the solution to equation

$$\mathrm{E}\left[Z_j\right] - c(a_j) = \mathrm{E}\left[\min\{g_j, Z_j\}\right]$$

and by defining $\hat{Z}_j := \min\{g_j, Z_j\}$, we have

$$\max\{y, \mathrm{E}\left[\max\{y, Z_j\}\right] - c(a_j)\} = \mathrm{E}\left[\max\{y, \hat{Z}_j\}\right]$$

which in turn allows us to argue that

$$\mathrm{E}\left[\max\{y, W^*_{\mathcal{M}}\}\right] \ge \mathrm{E}\left[\max\{y, \hat{Z}_j\}\right]$$

for all $y \in \mathbb{R}$ and $j \in [k]$.

Using the identity $\max(a, b) = -\min(-a, -b)$, this implies that for all $j \in [k]$, the random variable $(-\hat{Z}_j)$ second-order stochastically dominates the random variable $(-W^*_{\mathcal{M}})$. From Lemma 4.12, this allows us to obtain mappings $m_j : \mathrm{supp}(\hat{Z}_j) \mapsto \Delta(\mathrm{supp}(W^*_{\mathcal{M}}))$ such that:

1. For each $j \in [k]$, $W^*_{\mathcal{M}}$ can be obtain by sampling a $z \sim \hat{Z}_j$ and then sampling from $m_j(z)$.

2. For each $z \in \text{supp}(\hat{Z}_j)$, we have $\mathrm{E}[m_j(z)] \geq z$.

We can now define the amortized cost functions ρ^π and the cost sharing vectors b^π. Fix any deterministic commitment $\pi \in \mathcal{C}(\mathcal{M})$ and let $j = \pi(\sigma) \in [k]$ be the fixed action that π takes at state σ. For each state $s \in R_j$, we use $\pi|s \in \mathcal{C}(\mathcal{M}_s)$ to denote the commitment π after we transition to state s; note that this is also a deterministic commitment. By definition, each MDP $\mathcal{M}_s$ has horizon up to H and thus by the induction hypothesis, it admits a mapping $\rho^{\pi|\sigma}(\cdot)$ and a non-negative cost sharing vector $b^{\pi|\sigma}$, satisfying the properties of the lemma. Now, consider any trajectory $\tau \in \Upsilon_\pi$ and observe that $\tau = \{\sigma, \tau_s\}$ for some $s \in R_j$ and some $\tau_s \in \Upsilon_\pi(s)$. We define

$$\rho^\pi(\tau) := m_j(\min\{g_j, \rho^{\pi|s}(\tau_s)\})$$

and

$$b^\pi_{\sigma\tau} := \mathrm{E}[\rho(\tau)] - \mathrm{E}\left[\rho^{\pi|s}(\tau_s)\right]$$

and for all other $s \in S_\pi \setminus \{\sigma\}$ and $\tau \in \Upsilon_\pi(s)$, we use the same cost share $b^\pi_{s\tau} = b^{\pi|s}_{s\tau}$ that was used in $\mathcal{M}_s$. The proof of the three properties follows precisely the same steps as Lemma 4.10. □

From Lemma A.6, we immediately obtain the following counterpart of Theorem 4.9; the proof follows exactly the same steps as the proof of Theorem 4.9 that we presented in Section 4, and is thus omitted.

THEOREM A.7. *In any instance $\mathbb{I} = (\mathbb{M}, \mathcal{F})$ of matroid-max-CICS,* $\mathrm{OPT}(\mathbb{I}) \leq \mathrm{E}\left[\max_{S \in \mathcal{F}} \sum_{i \in S} W^*_{\mathcal{M}_i}\right]$.

A.4 Local Approximation. Finally, we note that our notion of local approximation seamlessly extends to the maximization setting by simply changing the inequality order. In particular:

DEFINITION A.8. (LOCAL APPROXIMATION FOR MAXIMIZATION SETTINGS) *Let $\mathcal{M}$ be any MDP and let $\alpha \in (0, 1]$. We say that a commitment $\pi \in \mathcal{C}(\mathcal{M})$ is an α-local approximation for $\mathcal{M}$ if $f_{\mathcal{M}^\pi}(\alpha y) \geq \alpha \cdot f_{\mathcal{M}}(y)$ for all $y \in \mathbb{R}$.*

Notice that in the maximization setting, we have an α-local approximations for $\alpha \in (0, 1]$. By following exactly the same steps as in the proof of Theorem 5.2, the following composition theorem is immediate:

THEOREM A.9. (COMPOSITION THEOREM FOR MAXIMIZATION SETTINGS) *Let $\mathbb{I} = (\mathbb{M}, \mathcal{F})$ be any instance of matroid-max-CICS, where each constituent MDP $\mathcal{M}_i$ admits an α-local approximation under some commitment $\pi_i \in \mathcal{C}(\mathcal{M}_i)$. Then,* $\mathrm{CG}(\mathbb{I}) \geq \alpha$.

By combining Theorem A.9 with the optimality of the Water Draining Index policy (Theorem A.4) and the upper bound of Theorem A.7, we obtain a way to efficiently approximate the optimal solution for matroid-max-CICS assuming that the underlying MDPs achieve good local approximation guarantees.

B The Combinatorial Setting

In this section, we describe how our entire framework extends to combinatorial settings beyond the case of matroids. We will state our results in full generality and distinguish between minimization and maximization whenever needed. The framework we consider is based on [Singla, 2017] and its followup [Gupta et al., 2019], where the authors develop a technique for combinatorial selection over Markov chains. In this section, we show that under local approximation, their results can be seamlessly extended to arbitrary MDPs.

DEFINITION B.1. (CICS) *A Costly Information with Combinatorial Inspection (CICS) instance $\mathbb{I}$ is defined with respect to a set of Costly Information MDPs $\{\mathcal{M}_i\}_{i=1}^n$, a feasibility constraint $\mathcal{F} \subseteq 2^{[n]}$ and a function $h : \mathcal{F} \mapsto \mathbb{R}$. At each step, an algorithm chooses one of the MDPs and advances it through one of its actions. The game terminates once the algorithm accepts a feasible[11] set $S \in \mathcal{F}$ of MDPs.*

[11] We assume that the algorithm will always ensure that this happens, i.e. when having already accepted a set of MDPs S, it will never accept another MDP i for which $S \cup \{i\} \cup A \notin \mathcal{F}$ for all $A \subseteq [n]$.

For a specific run of the algorithm, let A denote the set of all actions the algorithm took, $S \in \mathcal{F}$ denote the set of terminated MDPs and T be the corresponding set of terminals it accepted. Then, the total cost of the algorithm for this run in the minimization setting is given by

$$\sum_{t \in T} v(t) + h(S) + \sum_{a \in A} c(a)$$

and the total utility of the algorithm for the run in the maximization setting is given by

$$\sum_{t \in T} v(t) + h(S) - \sum_{a \in A} c(a).$$

The optimal adaptive policy in the minimization (resp. maximization) setting is the policy of minimum (resp. maximum) expected cost (resp. utility).

We note that there are two differences with respect to the matroid setting. The first is that $\mathcal{F}$ can now be any arbitrary set of constraints. In fact, we won't even have to assume that the set is downwards or upwards close, as long as the algorithms always ensure that they accept a feasible set of MDPs. The second is the addition of the function $h(\cdot)$; this is a function that encodes an extra cost (or reward, depending on the setting) that does not depend on the precise terminals of the MDPs that were selected, but only on the set of terminated MDPs.

The notions of amortization, water filling/draining, optimality curves, surrogate costs and local approximation apply to each MDP separately, and thus are independent of the underlying combinatorial setting. Therefore, all these definitions extend to the combinatorial setting without change. Our **first contribution** is to extend Theorem 4.9 beyond the matroid setting and prove that the performance of the optimal adaptive policy in any combinatorial setting is bounded by the surrogate costs of the underlying MDPs. We note that while this result was known in the single-selection setting, we are the first to prove it for the general combinatorial setting.

THEOREM B.2. *Let $\mathbb{I} = (\mathcal{M}_1, \cdots, \mathcal{M}_n, \mathcal{F}, h)$ be any instance of CICS. For each $i \in [n]$, let $W^*_{\mathcal{M}_i}$ be the water filling (resp. water draining) surrogate costs in the minimization (resp. maximization) setting. Then:*

1. *For the minimization setting,* $\mathrm{OPT}(\mathbb{I}) \geq \mathrm{E}\left[\min_{S \in \mathcal{F}}(\sum_{i \in S} W^*_{\mathcal{M}_i} + h(S))\right]$.
2. *For the maximization setting,* $\mathrm{OPT}(\mathbb{I}) \leq \mathrm{E}\left[\max_{S \in \mathcal{F}}(\sum_{i \in S} W^*_{\mathcal{M}_i} + h(S))\right]$.

Our **second contribution** is to show that local approximation continues to imply composition results even in the combinatorial setting. In particular, we prove the following extension of Theorem 5.2.

THEOREM B.3. *Let $\mathbb{I} = (\mathcal{M}_1, \cdots, \mathcal{M}_n, \mathcal{F}, h)$ be any instance of CICS. For each $i \in [n]$, let $W^*_{\mathcal{M}_i}$ be the water filling (resp. water draining) surrogate costs in the minimization (resp. maximization) setting. Finally, for each $i \in [n]$, let $\pi_i \in \mathcal{C}(\mathcal{M}_i)$ be some commitment that α-locally approximates $\mathcal{M}_i$. Then:*

1. *For the minimization setting,* $\mathrm{E}\left[\min_{S \in \mathcal{F}}(\sum_{i \in S} W^*_{\mathcal{M}_i^{\pi_i}} + h(S))\right] \leq \alpha \cdot \mathrm{E}\left[\min_{S \in \mathcal{F}}(\sum_{i \in S} W^*_{\mathcal{M}_i} + h(S))\right]$.
2. *For the maximization setting,* $\mathrm{E}\left[\max_{S \in \mathcal{F}}(\sum_{i \in S} W^*_{\mathcal{M}_i^{\pi_i}} + h(S))\right] \geq \alpha \cdot \mathrm{E}\left[\max_{S \in \mathcal{F}}(\sum_{i \in S} W^*_{\mathcal{M}_i} + h(S))\right]$.

By combining Theorem B.2 with Theorem B.3, and assuming that our MDPs admit local approximation guarantees, we are left with the task of optimizing over the CICS instance that is generated by the commitments; observe that this is now an instance over Markov chains. In the case of matroid constraints, we showed that we can always efficiently achieve this via the water filling/draining index policy. However, depending on the combinatorial constraint, efficient optimization might not be possible – consider for example the case where each MDP is a single terminal state corresponding to a set of elements and we need to select a minimum cost set cover.

The final key to the puzzle will be a way of efficiently approximating the optimal policy for a CICS instance $\mathbb{I} = (\mathcal{M}_1, \cdots, \mathcal{M}_n, \mathcal{F}, h)$ over Markov chains. This is precisely the setting that is considered by [Gupta et al., 2019]. In this work, it is shown that a sufficient condition to get an efficient β-approximation to the optimal policy for a CICS instance over Markov chains is for the underlying pair $(\mathcal{F}, h)$ to admit a β-approximate frugal algorithm. We defer the reader to [Gupta et al., 2019] for the full details. Here, we will just mention some examples of such combinatorial settings:

- Matching constraints in the maximization setting admit a 0.5-approximate frugal algorithm.
- Facility location constraints in the minimization setting admit a 1.861-approximate frugal algorithm.
- Set cover constraints in the minimization setting admit a $\min(f, \log n)$-approximate frugal algorithm where f is the maximum number of sets in which a ground element is present.

Combining everything, we obtain the following result for the combinatorial setting:

COROLLARY B.4. *Let $\mathbb{I} = (\mathcal{M}_1, \cdots, \mathcal{M}_n, \mathcal{F}, h)$ be any instance of CICS such that:*

1. *Each MDP $\mathcal{M}_i$ admits an α-local approximation.*
2. *The combinatorial setting $(\mathcal{F}, h)$ admits a β-frugal algorithm.*

Then, $\mathrm{CG}(\mathbb{I}) \leq \alpha \cdot \beta$ *for minimization, and* $\mathrm{CG}(\mathbb{I}) \geq \alpha \cdot \beta$ *for maximization.*

In other words, if (i) we can efficiently compute an α-local approximate commitment for each MDP and (ii) the combinatorial setting admits a frugal algorithm, then we can efficiently construct a committing policy that $(\alpha\beta)$-approximates the optimal (non-committing) policy. We are left with the task of proving our extended Theorems B.2 and B.3. It is not hard to see that both proofs are identical to their matroid counterparts and are thus omitted. In particular:

1. The proof of Theorem B.2 follows exactly the same steps as the proofs of Theorem 4.9 (for the minimization setting) and Theorem A.7 (for the maximization setting) that were presented in Section 4 and Section A respectively. In these proofs, we separately bounded the cost/utility contribution of each MDP by the corresponding surrogate costs/values and simply invoked the feasibility of the optimal adaptive policy at the end. Thus, the exact same proofs imply Theorem B.2 for any feasibility constraint $\mathcal{F}$ and any cost/reward function $h(\cdot)$.
2. The proof of Theorem B.3 follows exactly the same steps as the proofs of Theorem 5.2 (for the minimization setting) and Theorem A.9 that were presented in Section 4 and Section A respectively. In particular, none of these proofs used the fact that $\mathcal{F}$ is a matroid constraint at any point, and it is also straightforward to see that they immediately extend for any cost/reward function $h(\cdot)$.

C Challenges of Composing Local Conditions

In this section, we demonstrate the challenges of extending Whittle's condition to obtain bounds on the commitment gap. Fix any instance $\mathbb{I} = (\mathbb{M}, \mathcal{F})$ of CICS. Whittle's condition states that if the (local) commitments π_i are optimal for *all* local games $(\mathcal{M}_i, y)$, then $\Pi = (\pi_1, \cdots, \pi_n)$ is optimal for $\mathbb{I}$ and thus the commitment gap is 1. Naturally, this is a very strong local condition and it is desirable to relax it into an approximation guarantee, while still being able to compose it into a global bound.

A natural approach would be to try and extend this result to approximation guarantees over the local games. However, this does not work. Consider, for example, a minimization MDP $\mathcal{M}$ with two 0-cost actions: the first action leads to a deterministic value of 1, and the second uniformly leads to a stochastic value of 0 or 1. Committing to each of these actions leads to an expected cost of $\min(y, 1)$ and $1/2 \cdot \min(y, 1)$ in the local game respectively; the deterministic action achieves a 2-approximation for all local problems $(\mathcal{M}, y)$. Now consider a single-selection CICS with n copies of the above MDP. If we commit to the deterministic action in each MDP, our cost is deterministically 1, whereas taking the stochastic action in each costs us $1/2^n$ – an exponential gap!

A different approach would be to consider the negation of Whittle's condition: if a (local) commitment π_i is strictly suboptimal for all local game $(\mathcal{M}_i, y)$, then no optimal policy for the CICS instance $\mathbb{I}$ will commit to π_i. This could potentially allow us to rule out some of the commitment policies and simplify our instance. However, we show that this isn't true: there are instances of CICS where the commitment gap is 1 (and thus the optimal policy is a committing policy) and yet the actions it commits to are unambiguously suboptimal!

The Example. We consider an instance of single-selection min-CICS consisting of two MDPs $(\mathcal{M}_1, \mathcal{M}_2)$. The MDP $\mathcal{M}_1$ is a depth 1 Markov chain, consisting of a single action of cost 3 and resulting to a uniformly random terminal state of value 0 or 50. The MDP $\mathcal{M}_2$ is of height 2. On the first level, there are two actions a_1, a_2 of

cost $c(a_1) = c(a_2) = 0$. Taking action a_1 deterministically results to a terminal state of value 5. Taking action a_2 results to a terminal state of value 0 with probability $1/2$ and to a non-terminal state s with probability $1/2$. On state s, there are two more actions a_3, a_4 of cost $c(a_3) = c(a_4) = 0$. Taking a_3 results to a terminal of value 12 and taking a_4 results to a uniformly random terminal of value 0 or 50.

Notice that there are three committing policies in $\mathcal{C}(\mathcal{M}_2)$, corresponding to action sequences $\{a_1\}$, $\{a_2, a_3\}$ and $\{a_2, a_4\}$; we use π_1, π_{23} and π_{24} to denote these policies. As we have seen, each of these committing policies π implies a Markov chain $\mathcal{M}_2^\pi$ and an optimality curve for the local game $(\mathcal{M}_2^\pi, y)$, denoting the cost of the optimal policy for the local game as a function of the outside option $y \geq 0$. For this specific example, it is not hard to see that:

- $f_1(y) := f_{\mathcal{M}_2^{\pi_1}}(y) = \min\{y, 5\}$.
- $f_{23}(y) := f_{\mathcal{M}_2^{\pi_{23}}}(y) = 0.5 \cdot \min\{y, 12\}$.
- $f_{24}(y) := f_{\mathcal{M}_2^{\pi_{24}}}(y) = 0.25 \cdot \min\{y, 50\}$.

A pictorial representation of the two MDPs and the corresponding optimality curves for the three committing policies on $\mathcal{M}_2$ is shown in Figure 10.

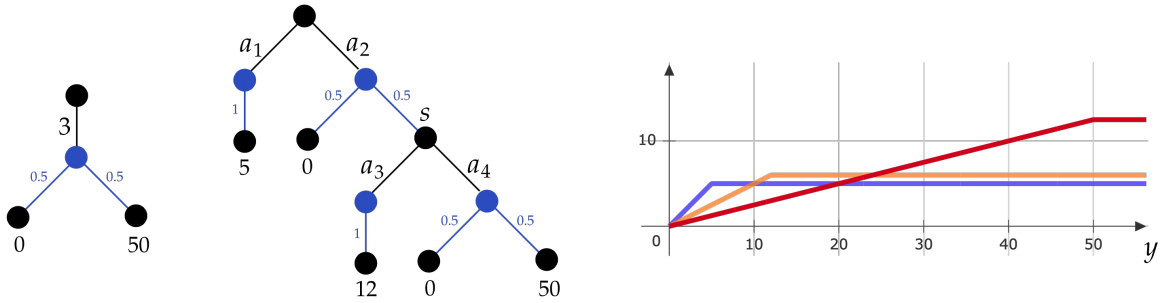


Figure 10: The MDPs $\mathcal{M}_1$ (left) and $\mathcal{M}_2$ (right). All the actions in $\mathcal{M}_2$ have a cost of 0. There are three committing policies for $\mathcal{M}_2$ with optimality curves $f_1(y)$ (blue), $f_{23}(y)$ (orange) and $f_{24}(y)$ (red).

The optimality curves demonstrate that for any value of y, the committing policy π_{23} is sub-optimal for the local game $(\mathcal{M}_2, y)$; in other words, the optimal policy for the local game will *never* take action a_3. One could expect that since a_3 is always suboptimal, the optimal policy would never take this action in the context of the bandit superprocess $(\mathcal{M}_1, \mathcal{M}_2)$. However, we can show that this isn't true. On the contrary, the optimal algorithm for $(\mathcal{M}_1, \mathcal{M}_2)$ will commit to π_{23}!

Claim C.1. *The optimal policy for the single-selection min-CICS instance $(\mathcal{M}_1, \mathcal{M}_2)$ is the committing policy that commits to the unique commitment for Markov chain $\mathcal{M}_1$ and to π_{23} for MDP $\mathcal{M}_2$.*

Proof. We prove the claim by computing the expected cost of all possible policies for the single-selection min-CICS instance $(\mathcal{M}_1, \mathcal{M}_2)$ and showing that the proposed committing policy has the minimum expected cost. At the beginning, every policy needs to either take the costly action of $\mathcal{M}_1$ or take either action a_1 or a_2 from $\mathcal{M}_2$.

- We first consider the set of policies that start by taking the costly action of $\mathcal{M}_1$ at a cost of 3; if the realized terminal has value 0 then we should clearly terminate, otherwise we play the local game on $\mathcal{M}_2$ for $y = 50$. In other words, the optimal policy that starts by taking the costly action of $\mathcal{M}_1$ has expected cost
$$3 + \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot \min\{f_1(50), f_{23}(50), f_{24}(50)\} = 5.5.$$
- Up next, we consider the set of policies that start by taking action a_1 in $\mathcal{M}_2$; in that case, we should play the local game on $\mathcal{M}_1$ with outside option $y = 5$ optimally. In other words, the optimal policy that starts by taking action a_1 on $\mathcal{M}_2$ has expected cost
$$\min\{5, 3 + \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 5\} = 5.$$

- Finally, we need to consider the set of policies that start by taking action a_2. If the realized outcome is the terminal state of value 0 (this happens with probability 1/2), then we should clearly accept it and terminate at a total cost of 0. Otherwise, we are left with the task of computing the optimal policy for the single-selection min-CICS $(\mathcal{M}_1, \mathcal{M}_s)$ where $\mathcal{M}_s$ is the sub-process of $\mathcal{M}_2$ rooted at state s. In other words, the cost of the optimal policy that starts by taking action a_2 on $\mathcal{M}_2$ will be precisely

$$\frac{1}{2} \cdot \mathrm{OPT}(\mathcal{M}_1, \mathcal{M}_s).$$

  We will now compute the cost of the optimal policy for the sub-instance $(\mathcal{M}_1, \mathcal{M}_s)$. Observe that any policy will either start by taking a_3 or a_4 or by taking the costly action in $\mathcal{M}_1$; in the latter case, if the outcome is 0 the policy should terminate, otherwise there is once again an optimal choice between a_3 and a_4. Thus, we can assume without loss that the optimal policy for the sub-instance $(\mathcal{M}_1, \mathcal{M}_s)$ will commit in advance to either a_3 or a_4. Furthermore, since both a_3 and a_4 have a cost of 0, we can assume that we start by taking the corresponding action. Thus:

  - If the policy commits to a_3, the optimal cost for the sub-instance $(\mathcal{M}_1, \mathcal{M}_s)$ is

$$\min\{12, 3 + \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot \min\{12, 50\}\} = 9.$$

  - If the policy commits to a_4, the optimal cost for the sub-instance $(\mathcal{M}_1, \mathcal{M}_s)$ is

$$\frac{1}{2} \cdot 0 + \frac{1}{2} \cdot \min\{50, 3 + \frac{1}{2} \cdot \min\{0, 50\} + \frac{1}{2} \cdot \min\{50, 50\}\} = 14.$$

  Thus, the cost of the optimal policy that starts by taking action a_2 on $\mathcal{M}_2$ will be $(1/2) \cdot 9 = 4.5$.

From the above case analysis, we obtain that the optimal policy for $(\mathcal{M}_1, \mathcal{M}_2)$ will start by taking action a_2, then action a_3 and finally the costly action of $\mathcal{M}_1$ - the policy terminates whenever a terminal state of value 0 is reached or it runs out of actions to take (in which case it accepts the terminal of value 12). Clearly, this is equivalent to the optimal policy that commits to π_{23}, concluding the proof. ∎

D Second Order Stochastic Dominance Lemma

In this section we provide a proof for Lemma 4.12. Once again, we note that this result is standard; here, we provide our own constructive proof for the sake of completeness and building intuition on how the water filling surrogate cost of an MDP is obtained. For simplicity, we refer to Lemma 4.12 as the Stochastic Dominance Lemma, henceforth SDL, which we restate for the reader's convenience.

LEMMA 4.12. (Second Order Stochastic Dominance.) *Let X, Z be discrete random variables that satisfy the property $\mathrm{E}[\min\{y, X\}] \leq \mathrm{E}[\min\{y, Z\}]$ for all $y \in \mathbb{R}$. There exists a mapping $m : \mathrm{supp}(Z) \mapsto \Delta(\mathrm{supp}(X))$ from the support of Z to distributions over the support of X such that:*

1. *X is obtained by sampling from $m(z)$ for a randomly sampled $z \sim Z$.*
2. *For all $z \in \mathrm{support}(Z)$, it holds that $\mathrm{E}[m(z)] \leq z$.*

Proof. We will say that $X \preceq Z$ if there exists a mapping $m : \mathrm{supp}(Z) \mapsto \Delta(\mathrm{supp}(X))$ such that the two conditions of the SDL hold. We associate each random variable W with a function $f_W(y) := \mathrm{E}[\min\{y, W\}]$ over $y \in \mathbb{R}$. In other words, we want to prove that

$$f_X(y) \leq f_Z(y) \ \ \forall y \in \mathbb{R} \Longrightarrow X \preceq Z.$$

We first prove **transitivity** of our condition. In other words, if X, Y, Z are discrete random variables such that $X \preceq Y$ and $Y \preceq Z$, we also have that $X \preceq Z$. The proof is immediate; let $m_1 : \mathrm{supp}(Z) \mapsto \Delta(\mathrm{supp}(Y))$ be the corresponding mapping for $Y \preceq Z$ and $m_2 : \mathrm{supp}(Y) \mapsto \Delta(\mathrm{supp}(X))$ be the corresponding mapping for $X \preceq Y$. Then, the composition mapping $m := m_2 \circ m_1$ that maps each $z \in \mathrm{supp}(Z)$ to a random realizations of

$m_2(y)$ for a randomly sampled $y \sim m_1(z)$ immediately satisfies both conditions and yields $X \preceq Z$. Formally, for each $x \in \text{supp}(X)$ we have

$$\Pr[X = x] = \sum_{y,z} \Pr[Z = z] \cdot \Pr[m_1(z) = y] \cdot \Pr[m_2(y) = x] = \sum_{z} \Pr[Z = z] \cdot \Pr[m(z) = x]$$

and for each $z \in \text{supp}(Z)$ we have

$$\mathrm{E}[m(z)] = \sum_{y} \Pr[m_1(z) = y] \cdot \mathrm{E}[m_2(y)] \leq \sum_{y} \Pr[m_1(z) = y] \cdot y = \mathrm{E}[m_1(z)] \leq z$$

and thus transitivity of the $\preceq$ operator is established.

We will now proceed to the **main proof**. Fix the random variables X and Z such that $f_X(y) \leq f_Z(y)$ for all $y \in \mathbb{R}$ and order their support sets so that

$$\mathcal{X} := \text{support}(X) = \{x_1, \cdots, x_M\}$$

and

$$\mathcal{Z} := \text{support}(Z) = \{z_1, \cdots, z_N\}$$

with $x_1 < x_2 < \cdots < x_M$ and $z_1 < z_2 < \cdots < z_N$. We also use $p_i^X := \Pr[X = x_i]$ for all $i \in [N]$ and $p_i^Z := \Pr[Z = Z_i]$ for all $i \in [M]$ to denote the corresponding probabilities, with

$$\sum_{i=1}^{M} p_i^X = \sum_{i=1}^{N} p_i^Z = 1.$$

We will say that X and Z **agree up to index i**, if $x_j = z_j$ and $p_j^X = p_j^Z$ for all $j < i$. Conventionally, we say that any two random variables will agree up to index 1 according to this definition. We will structure our proof of the SDL as an induction on the maximum index that X and Z agree up to. In particular, we break-down our proof in the following two steps:

1. (Induction Base). If X and Z agree up to index M, the SDL holds.
2. (Induction Step). If X and Z agree up to index $i \in [M-1]$, there exists a random variable Z' such that
   (a) Z' and X agree up to index $(i+1)$.
   (b) $Z' \preceq Z$.
   (c) For all $y \in \mathbb{R}$, $f_X(y) \leq f_{Z'}(y)$.

Before proving each of these two claims, let's see how they naturally construct an inductive proof for the SDL. Initially, we have that by definition, X and Z agree up to index 1. We can then apply our second claim (i.e. the induction step) to obtain a random variable $Z_1 \preceq Z$ that agrees with X up to index 2 and satisfies $f_X(y) \leq f_{Z_1}(y)$ for all $y \in \mathbb{R}$. We can then re-apply the induction step to obtain a random variable $Z_2 \preceq Z_1$ that agrees with X up to index 3 and satisfies $f_X(y) \leq f_{Z_2}(y)$ for all $y \in \mathbb{R}$. We keep applying the induction step for as long as we can, until we obtain a random variable $Z_{M-1} \preceq Z_{M-2} \preceq \cdots \preceq Z_1 \preceq Z$ that agrees with X up to index M and satisfies $f_X(y) \leq f_{Z_{M-1}}(y)$ for all $y \in \mathbb{R}$. We then proceed to use our first claim (i.e. the induction base) to show that $X \preceq Z_{M-1}$. Finally, we use the transitivity of operator $\preceq$ in order to obtain $X \preceq Z$, which concludes the proof.

Proof of Induction Base. Assume that X and Z agree up to index M; this means that $x_i = z_i$ and $p_i^X = p_i^Z$ for all $i < M$. Then, consider the deterministic mapping $m(z_i) = x_i$ if $i < M$ and $m(z_i) = x_M$ if $i \geq M$. Since x_M is the last point in the support $\mathcal{X}$, sampling $z \sim Z$ and outputting $m(z)$ is clearly equivalent to sampling $x \sim X$. For the expectation condition, we need to show that $m(z_i) = x_M \leq z_i$ for all $i \geq M$. For $y = x_M$, we have that $f_X(x_M) \leq f_Z(x_M)$. By definition:

$$f_X(x_M) = \sum_{i=1}^{M} p_i^X \cdot x_i = \sum_{i=1}^{M-1} p_i^X \cdot x_i + p_M^X \cdot x_M$$

and

$$f_Z(x_M) = \sum_{i=1}^{M-1} p_i^Z \cdot z_i + \sum_{i=M}^{N} p_i^Z \cdot \min(z_i, x_M).$$

Observe that the sums for $i \in [M-1]$ are equal by the agreement assumption and also $\sum_{i=M}^{N} p_i^Z = p_M^X$. Thus, to satisfy $f_X(x_M) \leq f_Z(x_M)$ we would need $x_M \leq \min(x_M, z_i)$ for all $i \geq M$ or equivalently that $x_M \leq z_M < z_{M+1} < \cdots < z_N$ and the proof follows.

Proof of Induction Step. We will now prove the induction step. Assume that the random variables X and Z agree up to index i for some $i \in [M-1]$; if they also agree up to index $(i+1)$ the step follows for $Z' = Z$ since clearly $Z \preceq Z$. If they don't agree up to index $(i+1)$, then by $f_X(y) \leq f_Z(y)$ for all $y \in \mathbb{R}$ it must necessarily be the case that either $x_i < z_i$ or ($x_i = z_i$ and $p_i^X > p_i^Z$). In any case, we deduce that $f_X(y) = f_Z(y)$ for all $y \leq x_i$ and $\lim_{y \to x_i^+} f_X(y) < \lim_{y \to x_i^+} f_Z(y)$.

Now, let $\ell(y)$ denote the unique straight line that passes through points $(x_i, f_X(x_i))$ and $(x_{i+1}, f_X(x_{i+1}))$, that is:

$$\ell(y) = \frac{f_X(x_{i+1}) - f_X(x_i)}{x_{i+1} - x_i} \cdot (y - x_i) + f_X(x_i).$$

Since $f_Z(y)$ is clearly concave, $\ell(y)$ will intersect it in at most two points; one of them is the point $(x_i, f_Z(x_i)) = (x_i, f_X(x_i))$ and the other point will necessarily be $(s, f_Z(s))$ for some $s \geq x_{i+1}$. A pictorial representation is given in Figure 11.

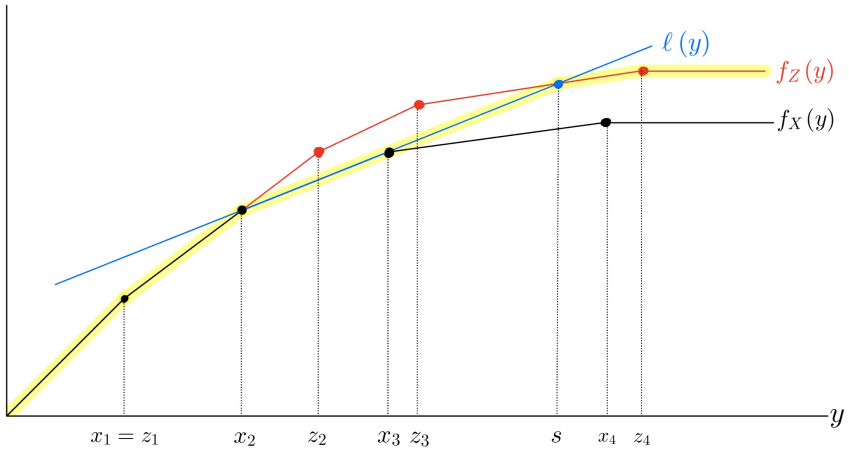


Figure 11: The induction step. Random variables X and Z agree up to index 2. The line $\ell(y)$ extends the line-segment of $f_X(y)$ from x_2 to x_3 and cuts $f_Z(y)$ on some $y = s \geq x_3$. Let $h(y)$ denote the highlighted curve. The random variable Z' that has curve $f_{Z'}(y) = h(y)$ agrees with X up to index 3 and satisfies $f_{Z'}(y) \geq f_X(y)$ for all $y \in \mathbb{R}$. Here, $J = \{2, 3\}$ denotes the set of indices of f_Z's break-points that lie in (x_2, s).

Now, consider the curve $h(y) = \min\{\ell(y), f_Z(y)\}$. By construction, this curve satisfies $f_X(y) \leq h(y)$ for all $y \in \mathbb{R}$. Furthermore, if it is the case that there exists some random variable Z' such that $f'_Z(y) = h(y)$, then it would be the case that Z' and X agree up to index $(i+1)$. Thus, to complete the proof, we need to show that such a random variable Z' not only exists, but also satisfies $Z' \preceq Z$. This will require it's own type of induction, so we will state it an prove it as a separate claim (Claim D.1) to ease the presentation. Notice that once this claim is proven, the proof of the SDL is completed. ∎

Claim D.1. *For any discrete random variable Z with curve $f_Z(y) = \mathrm{E}[\min\{y, Z\}]$ and any line $\ell(y)$ intersecting $f_Z(y)$ at exactly two points $a < b$, there always exists discrete random variable Z' such that $Z' \preceq Z$ and $f_{Z'}(y) = \min\{f_Z(y), \ell(y)\}$.*

Proof. Once again, we denote $\mathcal{Z} := \mathrm{support}(Z) = \{z_1, \cdots, z_N\}$ and assume $z_1 < z_2 < \cdots < z_N$. We also use $p_i^Z := \Pr[Z = z_i]$ for $i \in [N]$. Notice that the function $f_Z(y)$ is piece-wise linear, with breakpoints precisely at

$z_i \in \mathcal{Z}$. Furthermore, the slope at the interval $(-\infty, z_i]$ is 1, the slope at any interval $[z_i, z_{i+1}]$ for $i \in [N-1]$ is $1 - \sum_{j \le i} p_j^Z$ and the slope at the interval $[z_N, \infty)$ is 0. Finally, for each point $z_i \in \mathcal{Z}$, the difference of the slope of $f_Z(y)$ on the segment to its left minus the slope of $f_Z(y)$ on the segment to its right equals the probability p_i^Z.

We will begin by designing a useful **gadget**. This gadget $G(Z, a, b)$ takes as input a discrete random variable Z and two parameters $a < b$ such that there exists index $i \in [N]$ with $z_{i-1} \le a < z_i < b \le z_{i+1}$ (we denote $z_0 = -\infty$ and $z_{N+1} = +\infty$) and returns a random variable Z' that is obtained by mapping each point $z_j \in \mathcal{Z}$ with $j \neq i$ to itself, and mapping point z_i to point a with some probability λ and to point b with probability $1-\lambda$. Clearly, $\text{support}(Z') = \{a, b\} \cup \mathcal{Z} \setminus \{z_i\}$. Let $\ell_{ab}(y)$ be the line passing through points $(a, f_Z(a))$ and $(b, f_Z(b))$ and let s be the slope of this line. Furthermore, let s_1 be the slope of $f_Z(y)$ at the interval $[a, z_i]$ and s_2 be its slope at the interval $[z_i, b]$; by definition, $s_1 - s_2 = p_i^Z$. Also, note that $s_1 > s > s_2$ by concavity. Then, by using mapping probability

$$\lambda := \frac{s_1 - s}{s_1 - s_2}$$

it is not hard to see that $f_{Z'}(y) = \min\{f_Z(y), \ell_{ab}(y)\}$, since we maintain the probability mass at all points $z_j \neq z_i$ and furthermore we have $\Pr[Z' = a] = p_i^Z \cdot \lambda = s_1 - s$ and $\Pr[Z' = b] = p_i^Z \cdot (1-\lambda) = s - s_2$. Furthermore, it is also not hard to see that $Z' \preceq Z$; we only need to verify that

$$\lambda \cdot a + (1-\lambda) \cdot b \le z_i$$

or equivalently (by substituting λ's definition) that $s(b-a) \le s_1(z_i - a) + s_2(b - z_i)$; this always holds with equality due to the definition of s_1, s_2 and s.

We are now ready to prove the claim. Let $J = \{z_i \in (a, b)\}$; this is the set of points $z_i \in \mathcal{Z}$ for which $f_Z(z_i) > \ell(z_i)$. Notice that the gadget $G(Z, a, b)$ already proves the claim for the special case of $|J| = 1$. We will now inductively apply the gadget to prove the general case. Let z_i be the minimum point in J; we begin by applying the gadget $G(z_i, a, z_{i+1})$ to obtain a new random variable Z'; this is allowed since z_i is the unique point of $\mathcal{Z}$ in the interval (a, z_{i+1}). By our construction of the gadget and concavity of $f_Z(y)$, we have that $Z' \preceq Z$ and also that $f_{Z'}(y) \ge \min\{f_Z(y), \ell(y)\}$ for all $y \in \mathbb{R}$. Furthermore, the corresponding J-interval for Z' will now have one less point; thus, we can inductively keep applying our gadget and by transitivity of the $\preceq$ operator the claim follows. ∎

References

Ali Aouad, Jingwei Ji, and Yaron Shaposhnik. The pandora's box problem with sequential inspections. *Available at SSRN 3726167*, 2020.

Arash Asadpour and Hamid Nazerzadeh. Maximizing stochastic monotone submodular functions. *Management Science*, 62(8):2374–2391, 2016.

Ben Berger, Tomer Ezra, Michal Feldman, and Federico Fusco. Pandora's problem with combinatorial cost. *Proceedings of the 24th ACM Conference on Economics and Computation*, 2023.

Hedyeh Beyhaghi. Approximately-optimal mechanisms in auction design, search theory, and matching markets. *Cornell University*, 2019.

Hedyeh Beyhaghi and Linda Cai. Pandora's problem with nonobligatory inspection: Optimal structure and a PTAS. *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, 2022.

Hedyeh Beyhaghi and Linda Cai. Recent developments in pandora's box problem: Variants and applications. *SIGecom Exch.*, 21(1):20–34, October 2024.

Hedyeh Beyhaghi and Robert D. Kleinberg. Pandora's problem with nonobligatory inspection. *Proceedings of the 2019 ACM Conference on Economics and Computation*, 2019.

Shant Boodaghians, Federico Fusco, Philip Lazos, and Stefano Leonardi. Pandora's box problem with order constraints. *Proceedings of the 21st ACM Conference on Economics and Computation*, 2020.

Robin Bowers and Bo Waggoner. Matching with nested and bundled pandora boxes. *arXiv*, abs/2406.08711, 2024a.

Robin Bowers and Bo Waggoner. High-welfare matching markets via descending price. In *Web and Internet Economics: 19th International Conference, WINE 2023, Shanghai, China, December 4–8, 2023, Proceedings*, page 59–76, 2024b.

Robin Bowers, Elias Lindgren, and Bo Waggoner. Prophet inequalities for bandits, cabinets, and dags. *arXiv*, abs/2502.08976, 2025.

David Brown and James Smith. Optimal sequential exploration: Bandits, clairvoyants, and wildcats. *Operations Research*, 61:644–665, June 2013.

Shuchi Chawla, Evangelia Gergatsouli, Yifeng Teng, Christos Tzamos, and Ruimin Zhang. Pandora's box with correlations: Learning and approximation. *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, 2019.

Shuchi Chawla, Evangelia Gergatsouli, Jeremy McMahan, and Christos Tzamos. Approximating pandora's box with correlations. In *International Workshop and International Workshop on Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, 2021.

Shuchi Chawla, Dimitris Christou, and Trung Dang. Commitment gap via correlation gap. *arXiv preprint arXiv:2508.20246*, 2025.

Laura Doval. Whether or not to open pandora's box. *Journal of Economic Theory*, 175:127–158, 2018.

Ioana Dumitriu, Prasad Tetali, and Peter Winkler. On playing golf with two balls. *SIAM Journal on Discrete Mathematics*, 16(4):604–615, 2003.

Hossein Esfandiari, Mohammad Taghi Hajiaghayi, Brendan Lucier, and Michael Mitzenmacher. Online pandora's boxes and bandits. In *AAAI Conference on Artificial Intelligence*, 2019.

Hu Fu, Jia-Wen Li, and Daogao Liu. Pandora box problem with nonobligatory inspection: Hardness and approximation scheme. *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, 2022.

Hans Föllmer and Alexander Schied. *Stochastic Finance - An Introduction in Discrete Time*. De Gruyter, 2016.

Evangelia Gergatsouli and Christos Tzamos. Weitzman's rule for pandora's box with correlations. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, 2024.

John C Gittins. Bandit processes and dynamic allocation indices. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 41(2):148–164, 1979.

Kevin D. Glazebrook. On a sufficient condition for superprocesses due to whittle. *Journal of Applied Probability*, 19:99 – 110, 1982.

Sudipto Guha, Kamesh Munagala, and Saswati Sarkar. Information acquisition and exploitation in multichannel wireless networks. *arXiv*, abs/0804.1724, 2008.

Anupam Gupta, Haotian Jiang, Ziv Scully, and Sahil Singla. The Markovian Price of Information. In *Conference on Integer Programming and Combinatorial Optimization*, 2019.

Dylan Hadfield-Menell and Stuart J. Russell. Multitasking: Optimal planning for bandit superprocesses. In *Conference on Uncertainty in Artificial Intelligence*, 2015.

Martin Hoefer and Kevin Schewior. Threshold Testing and Semi-Online Prophet Inequalities. In *31st Annual European Symposium on Algorithms (ESA)*, 2023.

Martin Hoefer, Kevin Schewior, and Daniel Schmand. Stochastic probing with increasing precision. *SIAM Journal on Discrete Mathematics*, 38(1):148–169, 2024.

T. Ke and J. Villas-Boas. Optimal learning before choice. *Journal of Economic Theory*, 180, 2019.

Bobby Kleinberg, Bo Waggoner, and E. Glen Weyl. Descending price optimally coordinates search. *Proceedings of the 2016 ACM Conference on Economics and Computation*, 2016.

P. Nash. *Optimal Allocation of Resources Between Research Projects*. University of Cambridge, 1973.

Wojciech Olszewski and Richard R. Weber. A more general pandora rule? *Journal of Economic Theory*, 160: 429–437, 2015.

Ziv Scully and Laura Doval. Local hedging approximately solves pandora's box problems with optional inspection. *arXiv*, abs/2410.19011, 2024.

Sahil Singla. The price of information in combinatorial optimization. In *ACM-SIAM Symposium on Discrete Algorithms*, 2017.

V. Strassen. The Existence of Probability Measures with Given Marginals. *The Annals of Mathematical Statistics*, 36(2):423 – 439, 1965.

Richard Weber. On the Gittins Index for Multiarmed Bandits. *The Annals of Applied Probability*, 2(4):1024 – 1033, 1992.

Martin L Weitzman. Optimal search for the best alternative. *Econometrica*, pages 641–654, 1979.

Peter Whittle. Multi-Armed Bandits and the Gittins Index. *Journal of the Royal Statistical Society: Series B (Methodological)*, 42(2):143–149, 1980.