Cost-Aware Exploration for Structured Bandits

Xinyu Liu[1], Wei You[1], and Chao Qin[2]

[1]Department of Industrial Engineering and Decision Analytics, The Hong Kong University of Science and Technology, xliufn@connect.ust.hk, weiyou@ust.hk

[2]Stanford Graduate School of Business, Stanford University, chaoqin@stanford.edu

February 5, 2026

Abstract

We study pure exploration with heterogeneous per-measurement costs. An agent sequentially selects among K arms whose observation laws belong to (possibly structured) canonical exponential families, and each pull incurs an arm- and instance-dependent cost. The goal is to identify an instance-dependent answer (e.g., best arm, thresholding bandits, or Pareto set identification) while maximizing the rate at which the posterior probability of error decays *per unit spent budget*. We characterize the optimal cost-normalized posterior error exponent as the value of a maximin program that trades off statistical discrimination against the average cost, and show that no adaptive sampling rule can exceed this exponent. By working with the cost normalization, the exponent is characterized by a concave maximization problem. Motivated by this characterization, we develop a cost-aware pure exploration algorithm. The resulting dynamics is inherently nonsmooth and set-valued due to argmin operations, boundary effects, and normalization. We analyze the stochastic iterates through a continuous-time approximation based on differential inclusions and prove that, under mild regularity conditions, the algorithm attains the optimal cost-normalized posterior error exponent almost surely. Our results provide a general asymptotic optimality guarantee for cost-aware pure exploration beyond best-arm identification, covering broad exploration queries, exponential-family rewards, and structured bandit models, with substantially lower per-iteration computational overhead than optimize-then-track baselines such as Track-and-Stop.

1 Introduction

Pure exploration seeks to identify an instance-dependent answer (e.g., a best arm, a top-m set, or the arms above a threshold) from sequential noisy measurements. Classical formulations measure efficiency in the *number of samples*, effectively treating each pull as unit cost. In many applications, however, the binding constraint is a heterogeneous *budget* (time, money, compute, or human effort), and arms can consume that budget at vastly different rates. Such cost heterogeneity arises naturally in clinical screening (tests with different prices and turnaround times), online experimentation

(traffic and operational costs), sensing and robotics (energy), multi-fidelity simulation (cheap proxies vs. expensive high fidelity), and data-centric workflows such as prompt selection or human labeling (token/annotation costs). Information-theoretically, this shifts the objective from information per sample to *information per unit cost*. As a result, a sample-optimal policy can be budget-inefficient, over-investing in expensive measurements even when cheaper ones provide comparable discriminative power. The mismatch is particularly pronounced in structured bandits such as linear bandits, where geometry couples arms: a cheaper arm can be *more* informative because it probes directions that best separate the true instance from its hardest alternatives.

A growing literature incorporates sampling costs into pure exploration. Qin and Russo (2024) bridges regret minimization and pure exploration in a cost-aware setting for unstructured bandits. Related directions include multi-fidelity feedback (Poiani et al. 2024), LLM prompt selection (Hu et al. 2025), and other cost-sensitive designs (Kanarios et al. 2024, He-Yueya et al. 2025). Several works also pursue asymptotic optimality: Qin and Russo (2024) combine IDS with best-arm identification; Kanarios et al. (2024), Wu et al. (2025) adapt Track-and-Stop; and Poiani et al. (2024) extends Ménard (2019) and analyze subgradient-type schemes for multi-fidelity trade-offs.

In this work, we extend the algorithm of Qin and You (2025) to incorporate heterogeneous costs and general structured bandit models. Doing so is technically nontrivial for three reasons. First, pure-exploration designs rely on repeated arg min operations (e.g., to detect the most confusing alternative), which become set-valued under ties; existing analyses often address this via carefully constructed subdifferential subspaces (Wang et al. 2021). Second, structured bandits frequently admit sparse optimal allocations, pushing the dynamics toward the simplex boundary where Chernoff information can be nonsmooth and Clarke-type generalized gradients are required. Third, the information-directed selection (IDS) rule (You et al. 2023) we build upon involves a normalization whose denominator may vanish, creating genuine singularities. To handle these pathologies and establish convergence, we develop a continuous-time approximation based on *differential inclusions*. This framework naturally accommodates set-valued detection/selection rules and boundary-induced nonsmoothness, enabling a Lyapunov-based analysis of the resulting learning dynamics.

Contributions. We now summarize our contributions and compare with relevant literature.

First, we propose a simple cost-aware pure-exploration algorithm by extending Qin and You (2025) to cost-sensitive and structured settings. Many asymptotically optimal methods follow an "optimize-then-track" blueprint (Kanarios et al. 2024, Wu et al. 2025), solving a maximin program and tracking its solution online. Track-and-Stop (Garivier and Kaufmann 2016) is the canonical example, but its per-round allocation solve can be costly beyond small unstructured instances. Related approaches reduce computation via online mirror ascent (Ménard 2019) or game-theoretic no-regret dynamics (Degenne et al. 2019, 2020). Frank–Wolfe Sampling (Wang et al. 2021) offers a broadly applicable conditional-gradient tracker, but still requires a per-round linear subproblem and tuning of an r-subdifferential characterization. In contrast, our algorithm reduces per-iteration overhead and completely remove tracking steps by directly mapping the current state to sampling probabilities through the Chernoff-information structure.

Second, we provide the first asymptotic optimality guarantee for IDS beyond Gaussian best-arm identification. Our results cover general pure-exploration queries under one-parameter exponential-family rewards and structured bandit models, addressing the open questions raised in Qin and You (2025), Jourdan (2024). Top-two algorithms, for instance, sample via a β-biased randomization between a leader and challenger (Russo 2020, Qin et al. 2017, Shang et al. 2020, Jourdan et al. 2022). Despite strong empirical performance, optimality can require nontrivial tuning of β (Russo 2020). IDS (You et al. 2023) replaces the fixed-β rule with an information-gain criterion and yields asymptotic optimality for Gaussian BAI, while Qin and Russo (2024) proposes IDS for cost-aware BAI; until this work, guarantees beyond the Gaussian setting remained limited.

Third, we develop a proof technique tailored to the nonsmooth, set-valued nature of pure-exploration dynamics, offering a powerful toolkit for analyzing bandit algorithms that behave like nonsmooth first-order methods. We analyze the IDS updates through differential inclusions, leveraging nonsmooth chain rules (path differentiability) (Davis et al. 2020, Bolte and Pauwels 2021) and weak Sard-type properties (Benaïm et al. 2005) to construct Lyapunov arguments. Our analysis follows the stochastic-approximation perspective, which links discrete-time iterates to continuous-time limits via differential inclusions (Benaïm et al. 2005, Borkar 2008, Davis et al. 2020, Bolte and Pauwels 2021). While related tools are standard in nonsmooth optimization, they are not designed for the distinctive singularities induced by bandit sampling and IDS.

2 Problem Formulation

We study pure exploration with heterogeneous per-measurement costs. There are K arms, indexed by $[K] \triangleq \{1, \dots, K\}$. The unknown state of nature is represented by a parameter vector $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^d$. Sampling arm i produces a noisy measurement whose distribution we denote by $P_{\boldsymbol{\theta},i}$. Measurements are assumed independent across rounds, and conditionally identically distributed given the selected arm and $\boldsymbol{\theta}$. We assume that for each arm i, the observation law belongs to a one-dimensional canonical exponential family. Specifically, the law $P_{\boldsymbol{\theta},i}$ has density $p_{\boldsymbol{\theta},i}(y) = b(y)\exp\big(\eta_i(\boldsymbol{\theta})T(y) - A(\eta_i(\boldsymbol{\theta}))\big)$, where $\eta_i(\cdot)$ is the arm-specific natural-parameter map that allows for structured bandits.

Example 1 (Best-arm identification in linear bandits)**.** *Consider a linear bandit with unknown parameter $\boldsymbol{\theta} \in \mathbb{R}^d$ and design matrix $X \in \mathbb{R}^{K\times d}$, whose i-th row is the feature vector $\boldsymbol{x}_i^\top$. The mean reward of arm i is $m_i \triangleq \boldsymbol{x}_i^\top \boldsymbol{\theta}$. Pulling arm i yields an observation $Y \sim \mathcal{N}(m_i, \sigma_i^2)$, and observations are independent across pulls (and across arms). A canonical objective is best-arm identification: determine the (assumed unique) optimal arm $I^* \in \arg\max_{i\in[K]} m_i$.*

Assumption 1. *The parameter set $\Theta \subset \mathbb{R}^d$ is compact with nonempty interior, each η_i is continuous on Θ, and A is twice continuously differentiable on an open neighborhood of $\eta_i(\Theta)$ for each i. Moreover, there is a finite constant $L_A < \infty$ such that $\sup_{i\in[K]} \sup_{\boldsymbol{\theta}\in\Theta} |A'(\eta_i(\boldsymbol{\theta}))| \le L_A$.*

At each round $n \ge 1$, an $\mathcal{H}_{n-1}$-measurable sampling rule selects an arm $I_n \in [K]$ and observes $Y_n \sim P_{\boldsymbol{\theta},I_n}$, where $\mathcal{H}_n \triangleq \sigma(I_1, Y_1, \dots, I_n, Y_n)$ is the interaction history up to round n.

To capture cost heterogeneity across arms, we assume that sampling arm i under the true instance $\boldsymbol{\theta}$ incurs a (per-sample) cost $C_i(\boldsymbol{\theta})$.

Assumption 2. *The sampling costs are known continuous functions. Furthermore, there exist constants* $0 < c_{\min} \le c_{\max} < \infty$ *such that for all arms* i *and all* $\boldsymbol{\theta} \in \Theta$, $c_{\min} \le C_i(\boldsymbol{\theta}) \le c_{\max}$.

Let $\Delta_K \triangleq \{\boldsymbol{q} \in \mathbb{R}^K_{\ge 0} : \sum_{i=1}^K q_i = 1\}$ denote the simplex in $\mathbb{R}^K$. Define the sample allocation $\boldsymbol{p}_n \triangleq (p_{n,1}, \ldots, p_{n,K}) \in \Delta_K$, where $p_{n,i} \triangleq N_{n,i}/n$ and $N_{n,i} \triangleq \sum_{\ell=1}^n \mathbb{1}\{I_\ell = i\}$. For any $\boldsymbol{p} \in \Delta_K$, denote $\bar{C}_{\boldsymbol{\theta}}(\boldsymbol{p}) = \sum_{i=1}^K p_i C_i(\boldsymbol{\theta})$ as the average cost per sample. Then, the budget spent after n rounds is $B_n \triangleq \sum_{i=1}^K N_{n,i} C_i(\boldsymbol{\theta}) = n\bar{C}_{\boldsymbol{\theta}}(\boldsymbol{p}_n)$.

Posterior Error Exponent Per Unit Cost. The goal is to correctly identify an instance-dependent answer $\mathcal{I}(\boldsymbol{\theta})$ based on these noisy measurements. For a given instance $\boldsymbol{\theta}$, define the *alternative set* as the collection of instances yielding a different answer,

$$\mathrm{Alt}(\boldsymbol{\theta}) \triangleq \{\boldsymbol{\vartheta} \in \Theta : \mathcal{I}(\boldsymbol{\vartheta}) \neq \mathcal{I}(\boldsymbol{\theta})\},$$

as in Garivier and Kaufmann (2016), Wang et al. (2021), Qin and You (2025).

Assumption 3. *The true instance* $\boldsymbol{\theta} \in \Theta$ *has a unique answer* $\mathcal{I}(\boldsymbol{\theta})$. *Moreover, the answer is identifiable from the family of observation laws in the sense that for any* $\boldsymbol{\vartheta} \in \Theta$,

$$P_{\boldsymbol{\theta},i} = P_{\boldsymbol{\vartheta},i} \ \ \forall i \in [K] \quad \Longrightarrow \quad \mathcal{I}(\boldsymbol{\vartheta}) = \mathcal{I}(\boldsymbol{\theta}).$$

Furthermore, $\mathrm{Alt}(\boldsymbol{\theta})$ *is non-empty and* $\mathrm{dist}(\boldsymbol{\theta}, \mathrm{Alt}(\boldsymbol{\theta})) > 0$.

The form of the alternative set depends on the query to be answered. This assumption is automatically satisfied if $\mathrm{Alt}(\boldsymbol{\theta})$ is open, e.g., in best-arm identification. Moreover, it also covers cases where $\mathrm{Alt}(\boldsymbol{\theta})$ is not necessarily open, e.g., in thresholding bandit problem (Chen et al. 2014, Locatelli et al. 2016), but we require that the true parameter is not on the boundary.

Let Π_0 be a prior on Θ, and let $\Pi_n(\cdot) \triangleq \Pi(\cdot \mid \mathcal{H}_n)$ denote the posterior distribution after n rounds. Following Russo (2020), our goal is to design a sampling rule that drives rapid decay of the posterior mass $\Pi_n(\mathrm{Alt}(\boldsymbol{\theta}))$—the posterior probability of identifying an incorrect answer—while accounting for heterogeneous sampling costs across arms.

As discussed in Russo (2020), Qin et al. (2017), under a well-designed sampling rule the posterior mass $\Pi_n(\mathrm{Alt}(\boldsymbol{\theta}))$ typically decays exponentially in the *number of total samples*. With cost considerations, we instead normalize by the cumulative budget spent B_n and study the (cost-normalized) posterior error exponent $-\frac{1}{B_n}\log \Pi_n(\mathrm{Alt}(\boldsymbol{\theta}))$. Accordingly, our objective is to characterize and attain the optimal cost-aware large-deviation rate of posterior concentration.

Cost-normalized discriminative information rate. The large-deviation decay of $\Pi_n(\mathrm{Alt}(\boldsymbol{\theta}))$ under adaptive sampling is characterized by Russo (2020, Proposition 5) for the unstructured bandit in the non-cost-aware setting. Extension to our setting here is relatively straightforward, as we

detail in Appendix A. In particular, we show that, up to subexponential factors, the posterior mass of any open set $\tilde{\Theta} \subset \Theta$ decays exponentially at rate $\inf_{\boldsymbol{\vartheta} \in \tilde{\Theta}} \Gamma(\boldsymbol{p}_n; \boldsymbol{\vartheta})$, where $\Gamma(\boldsymbol{p}_n; \boldsymbol{\vartheta}) \triangleq \sum_{i=1}^K p_i \mathrm{KL}(P_{\boldsymbol{\theta},i} \| P_{\boldsymbol{\vartheta},i})$ is the discrimination information rate under allocation $\boldsymbol{p} \in \Delta_K$ and alternative instance $\boldsymbol{\vartheta} \in \Theta$. In the cost-aware setting, we normalize by the cumulative budget spent B_n and obtain the heuristic approximation

$$-\frac{1}{B_n} \log \Pi_n(\mathrm{Alt}(\boldsymbol{\theta})) = \frac{n}{B_n} \left(-\frac{1}{n} \log \Pi_n(\mathrm{Alt}(\boldsymbol{\theta})) \right) \approx \inf_{\boldsymbol{\vartheta} \in \mathrm{Alt}(\boldsymbol{\theta})} \frac{\Gamma(\boldsymbol{p}_n; \boldsymbol{\vartheta})}{\bar{C}_{\boldsymbol{\theta}}(\boldsymbol{p}_n)}.$$

See Theorem 1 and Proposition 3 for rigorous statements.

Motivated by this posterior-exponent characterization, we define a cost-normalized discrimination rate for a fixed sample allocation $\boldsymbol{p} \in \Delta_K$ and any alternative instance $\boldsymbol{\vartheta} \in \Theta$

$$\Gamma^{\mathrm{c}}(\boldsymbol{p}; \boldsymbol{\vartheta}) \triangleq \frac{\sum_{i=1}^K p_i \mathrm{KL}(P_{\boldsymbol{\theta},i} \| P_{\boldsymbol{\vartheta},i})}{\sum_{i=1}^K p_i C_i(\boldsymbol{\theta})} = \frac{\Gamma(\boldsymbol{p}; \boldsymbol{\vartheta})}{\bar{C}_{\boldsymbol{\theta}}(\boldsymbol{p})}. \tag{1}$$

This quantity measures the expected amount of discriminative information *per unit cost* against the alternative $\boldsymbol{\vartheta}$ when samples are allocated according to $\boldsymbol{p}$. A natural design criterion is to select an allocation that maximizes the discrimination rate against the worst-case alternative, leading to

$$\Gamma^* \triangleq \max_{\boldsymbol{p} \in \Delta_K} \inf_{\boldsymbol{\vartheta} \in \mathrm{Alt}(\boldsymbol{\theta})} \Gamma^{\mathrm{c}}(\boldsymbol{p}; \boldsymbol{\vartheta}). \tag{2}$$

Sufficient condition for optimality. We characterize the maximal achievable cost-normalized large-deviation rate of posterior concentration and provide a convenient sufficient condition under which an algorithm attains the optimal exponent. As in Russo (2020), we impose the following mild boundedness condition on the prior.

Assumption 4. *The prior Π_0 admits a density π_0 with respect to Lebesgue measure on Θ such that*

$$0 < \inf_{\boldsymbol{\theta} \in \Theta} \pi_0(\boldsymbol{\theta}) \ \leq \ \sup_{\boldsymbol{\theta} \in \Theta} \pi_0(\boldsymbol{\theta}) < \infty.$$

Theorem 1. *Assume Assumptions 1–4. Then, for any adaptive sampling rule,*

$$\limsup_{n \to \infty} -\frac{1}{B_n} \log \Pi_n(\mathrm{Alt}(\boldsymbol{\theta})) \ \leq \ \Gamma^* \qquad \mathbb{P}_{\boldsymbol{\theta}}\text{-}a.s.$$

Moreover, suppose an algorithm produces sample allocation $\boldsymbol{p}_n$ satisfying

$$\lim_{n \to \infty} \inf_{\boldsymbol{\vartheta} \in \mathrm{Alt}(\boldsymbol{\theta})} \Gamma^{\mathrm{c}}(\boldsymbol{p}_n; \boldsymbol{\vartheta}) = \Gamma^* \qquad \mathbb{P}_{\boldsymbol{\theta}}\text{-}a.s. \tag{3}$$

Then,

$$\lim_{n \to \infty} -\frac{1}{B_n} \log \Pi_n(\mathrm{Alt}(\boldsymbol{\theta})) = \Gamma^* \qquad \mathbb{P}_{\boldsymbol{\theta}}\text{-}a.s.$$

Theorem 1 is convenient in that it reduces asymptotic optimality to an almost sure *value convergence* property: it suffices to show that the algorithm drives the worst-case cost-normalized discrimination rate $\inf_{\boldsymbol{\vartheta} \in \mathrm{Alt}(\boldsymbol{\theta})} \Gamma^{\mathrm{c}}(\boldsymbol{p}_n; \boldsymbol{\vartheta})$ to Γ^*.

3 The Algorithm

In this section, we adapt the Pitfall-Adapted Nomination (PAN) algorithm of Qin and You (2025) to our cost-aware setting.

3.1 Decomposition of the Alternative Set

We impose the following mild structural condition on the alternative set, commonly adopted in the pure exploration literature, e.g., in Wang et al. (2021).

Assumption 5. *The alternative set* $\mathrm{Alt}(\boldsymbol{\theta})$ *is a finite union of convex sets. That is, there exists a finite index set* $\mathcal{X}(\boldsymbol{\theta})$ *and convex sets* $\{\mathrm{Alt}_x(\boldsymbol{\theta}) : x \in \mathcal{X}(\boldsymbol{\theta})\}$ *such that* $\mathrm{Alt}(\boldsymbol{\theta}) = \bigcup_{x\in\mathcal{X}(\boldsymbol{\theta})} \mathrm{Alt}_x(\boldsymbol{\theta})$.

The set $\mathcal{X}(\boldsymbol{\theta})$ captures the fundamental types of "confusing scenarios" under which an alternative instance produces an answer different from $\mathcal{I}(\boldsymbol{\theta})$; see Qin and You (2025, Appendix A) for comprehensive examples.

For each confusing scenario $x \in \mathcal{X}(\boldsymbol{\theta})$ and any sample allocation $\boldsymbol{p}$, the following quantity captures the cost-normalized discriminative information available to rule it out:

$$D_x(\boldsymbol{p};\boldsymbol{\theta}) \triangleq \inf_{\boldsymbol{\vartheta}\in\mathrm{Alt}_x(\boldsymbol{\theta})} \Gamma^{\mathrm{c}}(\boldsymbol{p};\boldsymbol{\vartheta}) = \frac{1}{\bar{C}_{\boldsymbol{\theta}}(\boldsymbol{p})} \inf_{\boldsymbol{\vartheta}\in\mathrm{Alt}_x(\boldsymbol{\theta})} \sum_{i=1}^{K} p_i \mathrm{KL}(P_{\boldsymbol{\theta},i} \| P_{\boldsymbol{\vartheta},i}),$$

This allows us to rewrite (2) as $\Gamma^* = \max_{\boldsymbol{p}\in\Delta_K} \min_{x\in\mathcal{X}(\boldsymbol{\theta})} D_x(\boldsymbol{p};\boldsymbol{\theta})$.

3.2 Cost-Aware PAN Algorithm

The PAN algorithm of Qin and You (2025), originally developed for non-cost-aware exploration in unstructured bandits, proceeds in three steps. At each round n, it performs: (i) *estimation*, producing an estimator $\boldsymbol{\theta}_n$; (ii) *detection*, identifying the most confusing scenario (the *pitfall*) in $\mathcal{X}(\boldsymbol{\theta}_n)$; and (iii) *selection*, allocating measurements across arms. We extend PAN to incorporate two additional features, cost-aware exploration and structured bandits, as summarized in Algorithm 1.[1]

Estimation rule. The problem instance $\boldsymbol{\theta}$ is unknown and must be learned from noisy bandit feedback. We assume access to an estimation oracle that, at each round, maps the current history to an instance estimate. In unstructured bandits, the empirical mean vector provides such an oracle, while in linear bandits a natural choice is the (regularized) least-squares estimator.

Assumption 6. *For each round* n*, given the history* $\mathcal{H}_n$*, an estimation routine returns an estimate* $\boldsymbol{\theta}_n = \boldsymbol{\theta}_n(\mathcal{H}_n) \in \Theta$ *of the true instance* $\boldsymbol{\theta}$*. If* $N_{n,i} \to \infty$ *for all* i*, then* $\boldsymbol{\theta}_n \to \boldsymbol{\theta}$ *almost surely.*

[1]The algorithm is *anytime* in the sense that the sampling rule itself does not depend on T.

Detection rule. The detection rule identifies the currently most confusing alternative scenario $x \in \mathcal{X}(\boldsymbol{\theta}_n)$. Recall that $D_x(\boldsymbol{p}_n; \boldsymbol{\theta}_n)$ denotes the plug-in estimate of the cost-normalized discriminative information available to rule out scenario x under the current sample allocations $\boldsymbol{p}_n$. Accordingly, we select the scenario with the smallest value of $D_x(\boldsymbol{p}_n; \boldsymbol{\theta}_n)$, i.e.,

$$x_n \in \arg\min_{x \in \mathcal{X}(\boldsymbol{\theta}_n)} D_x(\boldsymbol{p}_n; \boldsymbol{\theta}_n), \qquad \text{breaking ties arbitrarily.} \tag{4}$$

Although $D_x(\boldsymbol{p}; \boldsymbol{\theta})$ is defined via a cost normalization, the argmin in (4) is in fact independent of the cost functions $C_i(\cdot)$. Indeed, for any fixed $(\boldsymbol{p}, \boldsymbol{\theta})$ we have $\bar{C}_{\boldsymbol{\theta}}(\boldsymbol{p}) > 0$ by Assumption 2 and the factor $1/\bar{C}_{\boldsymbol{\theta}}(\boldsymbol{p})$ does not depend on x. Therefore,

$$\arg\min_{x \in \mathcal{X}(\boldsymbol{\theta})} D_x(\boldsymbol{p}; \boldsymbol{\theta}) = \arg\min_{x \in \mathcal{X}(\boldsymbol{\theta})} \inf_{\boldsymbol{\vartheta} \in \mathrm{Alt}_x(\boldsymbol{\theta})} \sum_{i=1}^{K} p_i \mathrm{KL}\left(P_{\boldsymbol{\theta},i} \| P_{\boldsymbol{\vartheta},i}\right). \tag{5}$$

This is expected: the detection step is purely a statistical discrimination task and is therefore independent of the cost of information collection. Similarly, for a given x, sample allocations $\boldsymbol{p}$ and any problem instance $\boldsymbol{\theta} \in \Theta$, the hardest instance[2] within Alt_x is also independent of the cost:

$$\boldsymbol{\vartheta}_x = \boldsymbol{\vartheta}_x(\boldsymbol{p}, \boldsymbol{\theta}) \in \arg\min_{\boldsymbol{\vartheta} \in \mathrm{Alt}_x(\boldsymbol{\theta})} \Gamma^{\mathrm{c}}(\boldsymbol{p}; \boldsymbol{\vartheta}) = \arg\min_{\boldsymbol{\vartheta} \in \mathrm{Alt}_x(\boldsymbol{\theta})} \Gamma(\boldsymbol{p}; \boldsymbol{\vartheta}).$$

Selection rule. We generalize the Information-Directed Selection (IDS) rule of Qin and You (2025) into the cost-aware IDS sampling probabilities

$$H_i^x(\boldsymbol{p}; \boldsymbol{\theta}) \triangleq \frac{p_i \mathrm{KL}\left(P_{\boldsymbol{\theta},i} \| P_{\boldsymbol{\vartheta}_x,i}\right) / C_i(\boldsymbol{\theta})}{\sum_{j=1}^{K} p_j \mathrm{KL}\left(P_{\boldsymbol{\theta},j} \| P_{\boldsymbol{\vartheta}_x,j}\right) / C_j(\boldsymbol{\theta})}, \quad i \in [K]. \tag{6}$$

Let $\boldsymbol{H}^x(\boldsymbol{p}; \boldsymbol{\theta}) \triangleq \left(H_1^x(\boldsymbol{p}; \boldsymbol{\theta}), \ldots, H_K^x(\boldsymbol{p}; \boldsymbol{\theta})\right) \in \Delta_K$.

Forced exploration. To guarantee sufficient exploration, we include forced-exploration blocks of length K that begin at times $n = Km^2$ for integers $m \geq 0$; over the subsequent K rounds, each arm is pulled exactly once. Between forced blocks, `PAN` follows the cost-aware IDS rule (6). Accordingly, the sample allocation $\boldsymbol{p}_n \in \Delta_K$ evolves as

$$\boldsymbol{p}_{n+1} = \boldsymbol{p}_n + \frac{1}{n+1}(\boldsymbol{e}_{I_{n+1}} - \boldsymbol{p}_n), \qquad I_{n+1} \begin{cases} \sim \boldsymbol{H}^{x_n}(\boldsymbol{p}_n; \boldsymbol{\theta}_n), & \text{if } \sqrt{\lfloor n/K \rfloor} \notin \mathbb{Z}, \\ = 1 + (n \bmod K), & \text{if } \sqrt{\lfloor n/K \rfloor} \in \mathbb{Z}. \end{cases} \tag{7}$$

Our main result is the optimality of Algorithm 1.

Theorem 2. *Under Assumptions 1–7, Algorithm 1 satisfies*

$$\lim_{n \to \infty} -\frac{1}{B_n} \log \Pi_n(\mathrm{Alt}(\boldsymbol{\theta})) = \Gamma^* \qquad \mathbb{P}_{\boldsymbol{\theta}}\text{-a.s.}$$

[2]We establish uniqueness of such $\boldsymbol{\vartheta}_x$ under mild conditions in Lemma 1.

Algorithm 1 Cost-Aware `PAN` Algorithm

Require: Horizon $T \in \mathbb{N}$, estimation routine `est`

Initialize: $N_i(0) \leftarrow 0$ for all $i \in [K]$; initialize an estimate $\boldsymbol{\theta}_0$; set $\boldsymbol{p}_0 \leftarrow (1/K)\mathbf{1}$
for $n = 0, 1, \ldots, T-1$ **do**
  $\boldsymbol{\theta}_n \leftarrow \texttt{est}(\mathcal{H}_n)$ ▷ Estimation
  if $\sqrt{\lceil n/K \rceil} \in \mathbb{Z}$ **then**
    $I_{n+1} \leftarrow 1 + (n \bmod K)$ ▷ Forced exploration
  else
    $x_n \in \arg\min_{x \in \mathcal{X}(\boldsymbol{\theta}_n)} D_x(\boldsymbol{p}_n; \boldsymbol{\theta}_n)$, breaking ties arbitrarily ▷ Detection
    Draw $I_{n+1} \sim \boldsymbol{H}^{x_n}(\boldsymbol{p}_n; \boldsymbol{\theta}_n)$ using (6) ▷ Selection
  end if
  Pull arm I_{n+1}, observe $Y_{n+1} \sim P_{\boldsymbol{\theta}, I_{n+1}}$, update $\boldsymbol{p}_n$ and $\mathcal{H}_{n+1} = \sigma(\mathcal{H}_n, I_{n+1}, Y_{n+1})$
end for
return $\mathcal{I}_T = \mathcal{I}(\boldsymbol{\theta}_T)$ where $\boldsymbol{\theta}_T \leftarrow \texttt{est}(\mathcal{H}_T)$

4 Dynamics of the Cost Allocations

The optimal cost-normalized exponent (2) is defined over sample allocations $\boldsymbol{p} \in \Delta_K$ through the fractions, which is not concave in $\boldsymbol{p}$. For the dynamical analysis of Algorithm 1, it is therefore convenient to reparameterize the allocation by *cost allocations*, which linearize the objective and yield a concave maximin problem.

4.1 Cost-Allocation Iterations

Fix the true instance $\boldsymbol{\theta}$. For any $\boldsymbol{p} \in \Delta_K$, define the associated *cost allocation*

$$w_i = w_i(\boldsymbol{p}) \triangleq \frac{p_i C_i(\boldsymbol{\theta})}{\bar{C}_{\boldsymbol{\theta}}(\boldsymbol{p})}, \qquad \bar{C}_{\boldsymbol{\theta}}(\boldsymbol{p}) \triangleq \sum_{j=1}^{K} p_j C_j(\boldsymbol{\theta}), \qquad i \in [K]. \tag{8}$$

Thus w_i is the fraction of the average per-sample cost attributed to arm i under $\boldsymbol{p}$. By Assumption 2, the map $\boldsymbol{p} \mapsto \boldsymbol{w}(\boldsymbol{p})$ is a bijection on Δ_K.[3] Under (8), the cost-normalized information rate (1) becomes linear in $\boldsymbol{w}$:

$$\tilde{\Gamma}^{\mathrm{c}}(\boldsymbol{w}; \boldsymbol{\vartheta}) \triangleq \sum_{i=1}^{K} w_i \frac{\mathrm{KL}(P_{\boldsymbol{\theta},i} \| P_{\boldsymbol{\vartheta},i})}{C_i(\boldsymbol{\theta})}. \tag{9}$$

Indeed, $\Gamma^{\mathrm{c}}(\boldsymbol{p}; \boldsymbol{\vartheta}) = \tilde{\Gamma}^{\mathrm{c}}(\boldsymbol{w}(\boldsymbol{p}); \boldsymbol{\vartheta})$ for all $\boldsymbol{p} \in \Delta_K$ and $\boldsymbol{\vartheta} \in \Theta$. Consequently, the optimal exponent admits the equivalent concave maximization $\Gamma^* = \max_{\boldsymbol{w} \in \Delta_K} \inf_{\boldsymbol{\vartheta} \in \mathrm{Alt}(\boldsymbol{\theta})} \tilde{\Gamma}^{\mathrm{c}}(\boldsymbol{w}; \boldsymbol{\vartheta})$. Under the decomposition in Assumption 5,

$$\Gamma^* = \max_{\boldsymbol{w} \in \Delta_K} \min_{x \in \mathcal{X}(\boldsymbol{\theta})} D_x^w(\boldsymbol{w}; \boldsymbol{\theta}), \quad \text{where} \quad D_x^w(\boldsymbol{w}; \boldsymbol{\theta}) \triangleq \inf_{\boldsymbol{\vartheta} \in \mathrm{Alt}_x(\boldsymbol{\theta})} \tilde{\Gamma}^{\mathrm{c}}(\boldsymbol{w}; \boldsymbol{\vartheta}). \tag{10}$$

The next lemma collects structural properties used throughout the sequel.

[3] The cost allocations are introduced for the analysis purpose and is not used by the algorithm. Consequently, we allow it to access the true costs to simplify exposition.

Lemma 1. *Fix $\boldsymbol{\theta}$ and $x \in \mathcal{X}(\boldsymbol{\theta})$. For $\boldsymbol{w} \in \mathbb{R}^K_{\geq 0}$, the map $\boldsymbol{w} \mapsto D^w_x(\boldsymbol{w};\boldsymbol{\theta})$ is nonnegative, continuous, concave and coordinatewise nondecreasing on Δ_K, and degree-one homogeneous: $D^w_x(\lambda\boldsymbol{w};\boldsymbol{\theta}) = \lambda D^w_x(\boldsymbol{w};\boldsymbol{\theta})$ for all $\lambda > 0$. For $\boldsymbol{w} \in \mathbb{R}^K_{>0}$, the minimizer $\boldsymbol{\vartheta}_x(\boldsymbol{w};\boldsymbol{\theta}) \triangleq \arg\min_{\boldsymbol{\vartheta}\in\mathrm{cl}(\mathrm{Alt}_x(\boldsymbol{\theta}))} \tilde{\Gamma}^c(\boldsymbol{w},\boldsymbol{\vartheta})$ is unique, $D^w_x(\cdot;\boldsymbol{\theta})$ is continuously differentiable and $[\nabla_{\boldsymbol{w}} D^w_x(\boldsymbol{w};\boldsymbol{\theta})]_i = \mathrm{KL}(P_{\boldsymbol{\theta},i} \| P_{\boldsymbol{\vartheta}_x(\boldsymbol{w};\boldsymbol{\theta}),i})/C_i(\boldsymbol{\theta})$. Moreover, there exists $d_x(\boldsymbol{\theta}) > 0$ such that $\sum_{i=1}^K [\nabla_{\boldsymbol{w}} D^w_x(\boldsymbol{w};\boldsymbol{\theta})]_i \geq d_x(\boldsymbol{\theta})$ for all $\boldsymbol{w} \in \mathrm{int}(\Delta_K)$, and hence $D^w_x(\boldsymbol{w};\boldsymbol{\theta}) = \langle \boldsymbol{w}, \nabla_{\boldsymbol{w}} D^w_x(\boldsymbol{w};\boldsymbol{\theta})\rangle > 0$ on $\mathrm{int}(\Delta_K)$.*

We impose a gradient bound that is satisfied by most pure-exploration tasks; see Wang et al. (2021, Lemma 1).

Assumption 7. *For all $x \in \mathcal{X}(\boldsymbol{\theta})$ and all differentiability points $\boldsymbol{w}$, the gradient $\nabla_{\boldsymbol{w}} D^w_x(\boldsymbol{w};\boldsymbol{\theta})$ exists and is uniformly bounded: there is $M = M(\boldsymbol{\theta}) < \infty$ such that $\|\nabla_{\boldsymbol{w}} D^w_x(\boldsymbol{w};\boldsymbol{\theta})\|_\infty \leq M$.*

Cost allocations along the trajectory. To analyze the asymptotic behavior of the cost allocations along the trajectory, we denote by $\boldsymbol{w}_n$ the cost allocation associated with $\boldsymbol{p}_n$ and *true* $\boldsymbol{\theta}$, i.e.,

$$w_{n,i} \triangleq \frac{p_{n,i} C_i(\boldsymbol{\theta})}{\bar{C}_{\boldsymbol{\theta}}(\boldsymbol{p}_n)} = \frac{N_{n,i} C_i(\boldsymbol{\theta})}{B_n}, \qquad i \in [K].$$

From the sampling process, $\boldsymbol{w}_n$ follows the exact update

$$\boldsymbol{w}_{n+1} = \boldsymbol{w}_n + \alpha_{n+1}(\boldsymbol{e}_{I_{n+1}} - \boldsymbol{w}_n), \quad \text{where } \alpha_{n+1} \triangleq \frac{C_{I_{n+1}}(\boldsymbol{\theta})}{B_{n+1}} \text{ for } I_{n+1} \text{ defined in (7).} \tag{11}$$

Under Assumption 2, the stepsize satisfies $\frac{c_{\min}}{c_{\max}(n+1)} \leq \alpha_{n+1} \leq \frac{c_{\max}}{c_{\min}(n+1)}$. This recursion is the budget analogue of the frequentist update $\boldsymbol{p}_{n+1} = \boldsymbol{p}_n + \frac{1}{n+1}(\boldsymbol{e}_{I_{n+1}} - \boldsymbol{p}_n)$.

In the $\boldsymbol{w}$-coordinate system, the detection and selection rules can be naturally related to their counterparts in the $\boldsymbol{p}$-coordinates through the cost-weighted transformation.

Detection rule in $\boldsymbol{w}$-coordinates. The detection step (4) in $\boldsymbol{w}$-coordinate becomes

$$\begin{aligned}
x_n \in \arg\min_{x\in\mathcal{X}(\boldsymbol{\theta}_n)} D_x(\boldsymbol{p}_n;\boldsymbol{\theta}_n) &= \arg\min_{x\in\mathcal{X}(\boldsymbol{\theta}_n)} \frac{1}{\bar{C}_{\boldsymbol{\theta}_n}(\boldsymbol{p}_n)} \inf_{\boldsymbol{\vartheta}\in\mathrm{Alt}_x(\boldsymbol{\theta}_n)} \sum_{i=1}^K p_{n,i} \mathrm{KL}\left(P_{\boldsymbol{\theta}_n,i} \| P_{\boldsymbol{\vartheta},i}\right) \\
&= \arg\min_{x\in\mathcal{X}(\boldsymbol{\theta}_n)} \frac{1}{\bar{C}_{\boldsymbol{\theta}}(\boldsymbol{p}_n)} \inf_{\boldsymbol{\vartheta}\in\mathrm{Alt}_x(\boldsymbol{\theta}_n)} \sum_{i=1}^K p_{n,i} \mathrm{KL}\left(P_{\boldsymbol{\theta}_n,i} \| P_{\boldsymbol{\vartheta},i}\right) \\
&= \arg\min_{x\in\mathcal{X}(\boldsymbol{\theta}_n)} \inf_{\boldsymbol{\vartheta}\in\mathrm{Alt}_x(\boldsymbol{\theta}_n)} \sum_{i=1}^K w_{n,i} \frac{\mathrm{KL}\left(P_{\boldsymbol{\theta}_n,i} \| P_{\boldsymbol{\vartheta},i}\right)}{C_i(\boldsymbol{\theta})}.
\end{aligned} \tag{12}$$

Although the explicit expression for the detection rule in $\boldsymbol{w}$-coordinates involves the cost terms $C_i(\boldsymbol{\theta})$, as implied by (5), the resulting most confusing answer x_n is actually independent of the cost. This stems from the fact that the detection step is fundamentally a statistical decision problem, aimed at identifying the most confusing alternative hypothesis based on how distinguishable the current estimate is from each alternative. The costs influence how we allocate resources across the arms, but they do not change the underlying task of distinguishing between hypotheses.

Selection rule in w-coordinates. IDS admits an elegant gradient representation in the $\boldsymbol{w}$-coordinates. For a fixed $x \in \mathcal{X}(\boldsymbol{\theta})$ and at a point $\boldsymbol{w}$ where $D_x^w(\cdot;\boldsymbol{\theta})$ is differentiable and positive, we define

$$\boldsymbol{h}^x(\boldsymbol{w};\boldsymbol{\theta}) \triangleq \frac{\boldsymbol{w} \circ \nabla_{\boldsymbol{w}} D_x^w(\boldsymbol{w};\boldsymbol{\theta})}{D_x^w(\boldsymbol{w};\boldsymbol{\theta})} \ \in \ \Delta_K. \tag{13}$$

Since $D_x^w(\cdot;\boldsymbol{\theta})$ is homogeneous of degree 1, Euler's identity yields $D_x^w(\boldsymbol{w};\boldsymbol{\theta}) = \langle \boldsymbol{w}, \nabla_{\boldsymbol{w}} D_x^w(\boldsymbol{w};\boldsymbol{\theta})\rangle$, confirming that (13) indeed defines a probability vector. Such IDS distribution (13) at round n, i.e. $h_i^{x_n}(\boldsymbol{w}_n;\boldsymbol{\theta}_n)$ can be computed explicitly using Lemma 1. It relates to the `PAN` algorithm 1 sampling distribution $H_i^{x_n}(\boldsymbol{p}_n;\boldsymbol{\theta}_n) \propto p_{n,i}\mathrm{KL}(P_{\boldsymbol{\theta}_n,i}\|P_{\vartheta_x,i})/C_i(\boldsymbol{\theta}_n)$ through the following identity:

$$\begin{aligned} h_i^{x_n}(\boldsymbol{w}_n;\boldsymbol{\theta}_n) &= \frac{w_{n,i}\mathrm{KL}(P_{\boldsymbol{\theta}_n,i}\|P_{\vartheta_x,i})/C_i(\boldsymbol{\theta}_n)}{\sum_{j=1}^K w_j \mathrm{KL}(P_{\boldsymbol{\theta},j}\|P_{\vartheta_x,j})/C_j(\boldsymbol{\theta}_n)} = \frac{\frac{C_i(\boldsymbol{\theta})}{C_{\boldsymbol{\theta}}} p_{n,i}\mathrm{KL}(P_{\boldsymbol{\theta}_n,i}\|P_{\vartheta_x,i})/C_i(\boldsymbol{\theta}_n)}{\sum_{j=1}^K \frac{C_j(\boldsymbol{\theta})}{C_{\boldsymbol{\theta}}} p_j \mathrm{KL}(P_{\boldsymbol{\theta},j}\|P_{\vartheta_x,j})/C_j(\boldsymbol{\theta}_n)} \\ &= \frac{C_i(\boldsymbol{\theta}) p_{n,i}\mathrm{KL}(P_{\boldsymbol{\theta}_n,i}\|P_{\vartheta_x,i})/C_i(\boldsymbol{\theta}_n)}{\sum_{j=1}^K C_j(\boldsymbol{\theta}) p_j \mathrm{KL}(P_{\boldsymbol{\theta},j}\|P_{\vartheta_x,j})/C_j(\boldsymbol{\theta}_n)} = \frac{C_i(\boldsymbol{\theta}) H_i^{x_n}(\boldsymbol{p}_n;\boldsymbol{\theta}_n)}{\sum_{j=1}^K C_j(\boldsymbol{\theta}) H_j^{x_n}(\boldsymbol{p}_n;\boldsymbol{\theta}_n)}. \end{aligned} \tag{14}$$

Thus, while $\boldsymbol{p}$ tracks sampling *frequencies*, the $\boldsymbol{w}$ captures the evolution of how the *budget* is distributed across arms, and the IDS rule has a particularly simple form in these coordinates.

Expected drift in cost coordinates Let $q_{n,i} \triangleq \mathbb{P}(I_{n+1} = i \mid \mathcal{H}_n)$ denote the actual sampling distribution at round n (including forced explorations). Define the expected stepsize

$$\widetilde{\alpha}_{n+1} \triangleq \mathbb{E}[\alpha_{n+1} \mid \mathcal{H}_n] = \sum_{i=1}^K q_{n,i} \frac{C_i(\boldsymbol{\theta})}{B_n + C_i(\boldsymbol{\theta})},$$

and introduce the cost-weighted sampling distribution $\widetilde{\boldsymbol{h}}_n \in \Delta_K$ such that $\mathbb{E}[\alpha_{n+1}\boldsymbol{e}_{I_{n+1}}|\mathcal{H}_n] = \widetilde{\alpha}_{n+1}\widetilde{\boldsymbol{h}}_n$. Explicitly, the components of $\widetilde{\boldsymbol{h}}_n$ are given by

$$\widetilde{h}_{n,i} \triangleq q_{n,i} \frac{C_i(\boldsymbol{\theta})}{B_n + C_i(\boldsymbol{\theta})} \Big/ \Big[\sum_{j=1}^K q_{n,j} \frac{C_j(\boldsymbol{\theta})}{B_n + C_j(\boldsymbol{\theta})}\Big].$$

Taking conditional expectation in the update rule (11) yields the expected drift of $\boldsymbol{w}_n$:

$$\mathbb{E}\left[\boldsymbol{w}_{n+1} - \boldsymbol{w}_n \mid \mathcal{H}_n\right] = \widetilde{\alpha}_{n+1}(\widetilde{\boldsymbol{h}}_n - \boldsymbol{w}_n). \tag{15}$$

The following lemma characterizes the asymptotic relationship between the actual sampling distribution $\widetilde{\boldsymbol{h}}_n$ and the ideal IDS distribution $\boldsymbol{h}^{x_n}(\boldsymbol{w}_n;\boldsymbol{\theta}_n)$.

Lemma 2. *With estimator sequence $\boldsymbol{\theta}_n \to \boldsymbol{\theta}$ almost surely,*

1. ***Detection consistency:*** $\lim_{n\to\infty} \left|D_{x_n}^w(\boldsymbol{w}_n,\boldsymbol{\theta}_n) - \min_{x\in\mathcal{X}} D_x^w(\boldsymbol{w}_n,\boldsymbol{\theta}_n)\right| = 0$ *almost surely.*
2. ***Selection consistency:*** *On non-forced rounds, the sampling distribution $\widetilde{\boldsymbol{h}}_n$ satisfies*

$$|\widetilde{h}_{n,i} - h_i^{x_n}(\boldsymbol{w}_n,\boldsymbol{\theta}_n)| \le \frac{c_{\max}}{n c_{\min}} h_i^{x_n}(\boldsymbol{w}_n,\boldsymbol{\theta}_n) \quad \text{for all } i \in [K], \tag{16}$$

and consequently $\|\widetilde{\boldsymbol{h}}_n - \boldsymbol{h}^{x_n}(\boldsymbol{w}_n,\boldsymbol{\theta}_n)\|_\infty \to 0$ at rate $O(1/n)$.

Lemma 2 shows that, in cost coordinates, the stochastic sampling process asymptotically follows a deterministic drift direction given by the IDS distribution $\boldsymbol{h}^{x_n}(\boldsymbol{w}_n,\boldsymbol{\theta}_n)$. This gradient-like form of IDS naturally provides a convenient optimization perspective for our analysis.

4.2 Continuous-Time Dynamics

We study the discrete-time iterates via their limiting continuous-time dynamics as $n \to \infty$, i.e., as the step size vanishes. Throughout this subsection, we fix the true instance $\boldsymbol{\theta}$ and suppress it in the notation: write $D_x^w(\boldsymbol{w}) \equiv D_x^w(\boldsymbol{w}; \boldsymbol{\theta})$ and $\mathcal{X} \equiv \mathcal{X}(\boldsymbol{\theta})$.

Recall that the discrete IDS rule in (13) uses $\nabla_{\boldsymbol{w}} D_x^w$ and divides by D_x^w. Two issues therefore arise on simplex boundary: (i) D_x^w may fail to be differentiable, and (ii) the denominator can vanish. To isolate points where the discrete rule is well defined, define the *regular region*[4]

$$\mathcal{R} \triangleq \Big\{ \boldsymbol{w} \in \Delta_K : \min_{x \in \mathcal{X}} D_x^w(\boldsymbol{w}) > 0 \text{ and } D_x^w \text{ is } C^1 \text{ at } \boldsymbol{w}, \forall x \in \arg\min_{x' \in \mathcal{X}} D_{x'}^w(\boldsymbol{w}) \Big\}.$$

For each $\boldsymbol{w} \in \mathcal{R}$, the IDS sampling distribution is well defined. By Lemma 1, for every $x \in \mathcal{X}$ and $\boldsymbol{p} \in \mathrm{int}(\Delta_K)$, the function $D_x^w(\cdot; \boldsymbol{\theta})$ is strictly positive and C^1; hence $\mathrm{int}(\Delta_K) \subseteq \mathcal{R}$ and $\overline{\mathcal{R}} = \Delta_K$. Therefore IDS can be extended from $\mathcal{R}$ to the full simplex via an appropriate limit construction.

Defining a continuous-time analogue of IDS entails an additional subtlety. As $n \to \infty$, the detected index x_n can switch on a fast time scale, while the sample allocation $\boldsymbol{w}_n$ evolves on a slower time scale. Consequently, in the continuous-time limit, the sampling decision at time t should respond not to a single scenario, but to the entire *active set* $\arg\min_{x \in \mathcal{X}} D_x^w(\boldsymbol{w}(t))$. To capture the limiting effect of rapid switching among active scenarios, we introduce an *IDS correspondence*: a set-valued extension of the discrete IDS rule that collects all limiting sampling responses induced by such fast switching. These limiting responses are parameterized by an "average" detection-frequency vector $\boldsymbol{\nu} \in \mathtt{det}(\boldsymbol{w})$, where $\mathtt{det}(\boldsymbol{w})$ is the detection correspondence defined as

$$\mathtt{det}(\boldsymbol{w}) \triangleq \arg\min_{\boldsymbol{\nu} \in \Delta_{|\mathcal{X}|}} F(\boldsymbol{w}, \boldsymbol{\nu}).$$

At the slower time scale, the discrete IDS updates track a *smoothed* response to an average frequency of confusing scenarios, represented by some $\boldsymbol{\nu} \in \mathtt{det}(\boldsymbol{w})$. Define

$$F(\boldsymbol{w}, \boldsymbol{\nu}) \triangleq \sum_{x \in \mathcal{X}} \nu_x D_x^w(\boldsymbol{w}), \quad \forall \boldsymbol{w} \in \Delta_K, \quad \forall \boldsymbol{\nu} \in \Delta_{|\mathcal{X}|}.$$

For $\boldsymbol{w} \in \mathcal{R}$ and $\boldsymbol{\nu} \in \mathtt{det}(\boldsymbol{w})$, define the averaged IDS response

$$\boldsymbol{h}(\boldsymbol{w}, \boldsymbol{\nu}) \triangleq \sum_{x \in \mathcal{X}} \nu_x \boldsymbol{h}^x(\boldsymbol{w}) = \frac{\boldsymbol{w} \circ \nabla_{\boldsymbol{w}} F(\boldsymbol{w}, \boldsymbol{\nu})}{F(\boldsymbol{w}, \boldsymbol{\nu})} \ \in \ \Delta_K, \tag{17}$$

and the IDS selection correspondence

$$\mathtt{sel}(\boldsymbol{w}) \triangleq \mathrm{conv}\Big\{ \lim_{n \to \infty} \boldsymbol{h}(\boldsymbol{w}^n, \boldsymbol{\nu}^n) : \ \boldsymbol{w}^n \to \boldsymbol{w}, \ \boldsymbol{w}^n \in \mathcal{R}, \ \boldsymbol{\nu}^n \in \mathtt{det}(\boldsymbol{w}^n) \Big\}.$$

The continuous-time evolution of $\boldsymbol{w}$ is then described by the differential inclusion (DI)

$$\dot{\boldsymbol{w}}(t) \in G\big(\boldsymbol{w}(t)\big) \triangleq \mathtt{sel}\big(\boldsymbol{w}(t)\big) - \boldsymbol{w}(t), \qquad \boldsymbol{w}(0) = \boldsymbol{w}_0 \in \Delta_K. \tag{18}$$

It is worth noting the similarity between the DI and its discrete version in (15).

[4]For boundary $\boldsymbol{w}$, we interpret continuous differentiablity as C^1 on a neighborhood of $\boldsymbol{w}$ within the relative interior of the smallest face containing $\boldsymbol{w}$.

Proposition 1 (Properties of the DI)**.** *The correspondences* `det` *and* `sel` *are nonempty, convex-valued, compact-valued, and upper hemicontinuous on* Δ_K*. For every* $\boldsymbol{w}_0 \in \Delta_K$*, the differential inclusion* (18) *admits an absolutely continuous solution* $\boldsymbol{w}(\cdot)$ *with* $\boldsymbol{w}(0) = \boldsymbol{w}_0$*, and* Δ_K *is forward invariant. Moreover, for any solution* $\boldsymbol{w}(\cdot)$ *there exists a measurable selection* $t \mapsto \boldsymbol{s}(t) \in \texttt{sel}(\boldsymbol{w}(t))$ *such that* $\dot{\boldsymbol{w}}(t) = \boldsymbol{s}(t) - \boldsymbol{w}(t)$ *for a.e.* $t \geq 0$*.*

5 Convergence Analysis

The main result of this paper is the *value convergence* of Algorithm 1. We show that both the cost allocation and the sampling frequencies induce the optimal value.

Theorem 3. *Under Assumptions 1–7, let* $\{\boldsymbol{w}_n\}_{n\geq 0}$ *be the cost allocation sequence generated by Algorithm 1 and iteration* (11)*. Then,* $\min_{x\in\mathcal{X}} D_x^w(\boldsymbol{w}_n; \boldsymbol{\theta}_n) \to \Gamma^*$ *almost surely. As a consequence, for* $\{\boldsymbol{p}_n\}_{n\geq 0}$ *from Algorithm 1, we also have* $\min_{x\in\mathcal{X}} D_x(\boldsymbol{p}_n; \boldsymbol{\theta}_n) \to \Gamma^*$ *almost surely.*

In view of the sufficient condition in Theorem 1, the value convergence of the sampling allocations generated by Algorithm 1 (Theorem 3) certifies the asymptotic optimality claimed in Theorem 2. Our analysis proceeds in four steps.

First, in Appendix E.1, we construct a globally well-defined information value that serves as a Lyapunov function for the limiting differential inclusion and show that it is nondecreasing along every solution trajectory. By LaSalle's invariance principle, the associated ω-limit set has empty interior. Second, in Appendix E.2, we transfer this continuous-time structure to the stochastic iterates via stochastic-approximation tools—the asymptotic pseudo-trajectory (APT) framework and the characterization of internally chain transitive (ICT) sets—which yields convergence of the information value along the discrete-time trajectory. Third, in Appendix E.3, we rule out convergence to the degenerate value 0 under interior initialization, ensuring the dynamics remains in a positive-information regime. Finally, in Appendix E.4, we use the Kullback–Leibler divergence as an energy function on the active set and establish a uniform negative-drift bound. This forces the limiting information value to coincide with the optimal value F^*, completing the proof.

6 Numerical Experiment

We implement Algorithm 1 and compare against four baselines: (i) *Track-and-Stop*, where the sampling proportions are obtained by running the deterministic version (i.e., remove the estimation step and assume that $\widehat{\boldsymbol{\theta}}_n$ is the ground truth) of Algorithm 1 for 100 iterations on the current estimate $\widehat{\boldsymbol{\theta}}_n$, coupled with a classical track-and-stop rule; (ii) *LinGapE* (Xu et al. 2018); (iii) *OD-LinBAI* (Yang and Tan 2022); and (iv) *Uniform* sampling. All methods use a ridge-regression estimation oracle. Additional experiments are deferred to Appendix G (heterogeneous costs), and we refer to Qin and You (2025) for unit-cost variants in the fixed-confidence setting.

We consider fixed-confidence linear best-arm identification with $d = 5$ and $K = 6$. Arms $a_i = e_i$ for $i \in \{1, \ldots, 5\}$, and $a_6 = \cos(\omega)e_1 + \sin(\omega)e_2$ with $\omega = 0.01$, so a_6 is nearly collinear with a_1.

The true parameter is $\boldsymbol{\theta} = (2, 0, 0, 0, 0)$ with unit noise variance and unit costs, yielding $\mu_1 = 2$ and $\mu_6 = 2\cos(\omega) \approx 1.9999$. This is the classical hard instance of Soare et al. (2014): distinguishing a_1 from a_6 requires shrinking uncertainty along the nearly $-e_2$ contrast direction, so accurately estimating θ_2 is essential even though arm 2 is far from optimal in mean.

For each replicate, we record the smallest budget n such that the posterior error $\Pi_n(\mathrm{Alt}(\boldsymbol{\theta})) \leq 10^{-3}$, censoring at $B = 10^6$ if the target is not reached. Figure 1 reports the distribution of budgets needed to reach the target over 1000 replicates. Algorithm 1 is the best-performing method, with a median budget of 2.77×10^4, while Track-and-Stop is slightly slower with median 3.32×10^4. In contrast, LinGapE, OD-LinBAI, and Uniform have markedly larger medians (8.85×10^4, 1.08×10^5, and 1.44×10^5, respectively).

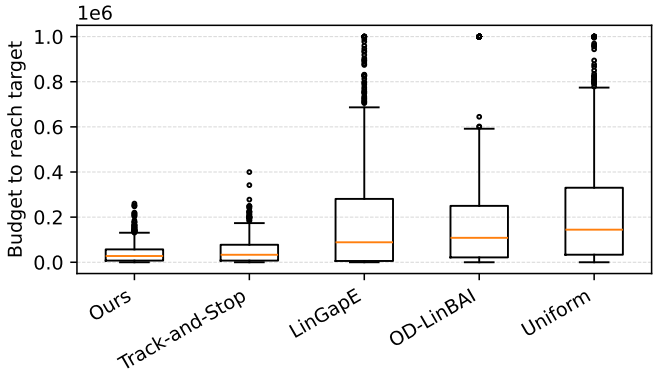

Figure 1: Budget to reach target posterior error $\Pi_n(\mathrm{Alt}(\theta)) \leq 10^{-3}$ over 1000 replicates.

The behavior matches the oracle structure: the optimal allocation places almost all mass on arm 2 ($p_2^* \approx 0.995$). Algorithm 1 quickly discovers and tracks this, with average final allocation ≈ 0.994 on arm 2. Track-and-Stop (with our approximate oracle) assigns less to arm 2 on average (≈ 0.756), while LinGapE, OD-LinBAI, and Uniform overweight the contender arms (1 and 6), slowing contraction along the critical contrast and producing heavy-tailed budgets.

7 Conclusion

We studied pure exploration with heterogeneous per-measurement costs and characterized the optimal large-deviation rate at which the posterior probability of error can decay *per unit spent budget*. Our main contribution is a cost-aware extension of the Pitfall-Adapted Nomination (PAN) framework that attains this optimal cost-normalized posterior error exponent almost surely under mild regularity conditions, for general one-parameter exponential-family rewards and structured bandit models. Conceptually, the results clarify that cost heterogeneity fundamentally changes the design criterion from "information per sample" to *information per unit cost*, and that asymptotically optimal behavior is governed by budget allocations rather than sampling frequencies.

Technically, we developed a proof strategy based on continuous-time approximation via differential inclusions, which accommodates fast switching in the detection step, nonsmoothness of the Chernoff

information objective, and the set-valued nature of IDS at ties and on the boundary of the simplex. This perspective may be useful beyond the present setting, offering a general toolkit for analyzing bandit sampling rules that behave like nonsmooth first-order methods.

Finally, our analysis incorporates forced exploration to streamline technical arguments and ensure sufficient excitation. Empirically, however, we observe that forced exploration is not essential for practical performance. Establishing asymptotic optimality without this mechanism is an interesting direction for future work.

References

Jean-Pierre Aubin and Arrigo Cellina. *Differential Inclusions: Set-Valued Maps and Viability Theory*. Springer Berlin Heidelberg, 1984.

Brian Beavis and Ian M Dobbs. *Optimization and stability theory for economic analysis*. Cambridge university press, 1990.

Michel Benaïm, Josef Hofbauer, and Sylvain Sorin. Stochastic approximations and differential inclusions. *SIAM Journal on Control and Optimization*, 44(1):328–348, 2005.

Pascal Bianchi, Walid Hachem, and Sholom Schechtman. Stochastic subgradient descent escapes active strict saddles on weakly convex functions. *Mathematics of Operations Research*, 49(3):1761–1790, 2024.

Jérôme Bolte and Edouard Pauwels. Conservative set valued fields, automatic differentiation, stochastic gradient methods and deep learning. *Mathematical Programming*, 188(1):19–51, 2021.

Vivek S Borkar. *Stochastic approximation: a dynamical systems viewpoint*, volume 9. Springer, 2008.

Haim Brezis. *Operateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*, volume 5. Elsevier, 1973.

Shouyuan Chen, Tian Lin, Irwin King, Michael R Lyu, and Wei Chen. Combinatorial pure exploration of multi-armed bandits. *Advances in neural information processing systems*, 27, 2014.

Damek Davis, Dmitriy Drusvyatskiy, Sham Kakade, and Jason D Lee. Stochastic subgradient method converges on tame functions. *Foundations of computational mathematics*, 20(1):119–154, 2020.

Rémy Degenne and Wouter M. Koolen. Pure exploration with multiple correct answers. *Advances in Neural Information Processing Systems*, 32, 2019.

Rémy Degenne, Wouter M. Koolen, and Pierre Ménard. Non-asymptotic pure exploration by solving games. *Advances in Neural Information Processing Systems*, 32, 2019.

Rémy Degenne, Pierre Ménard, Xuedong Shang, and Michal Valko. Gamification of pure exploration for linear bandits. In *International Conference on Machine Learning*, pages 2432–2442. PMLR, 2020.

Ignacio Esponda, Demian Pouzo, and Yuichi Yamamoto. Corrigendum to "Asymptotic behavior of Bayesian learners with misspecified models"[J. Econ. Theory 195 (2021) 105260]. *Journal of Economic Theory*, 204:105513, 2022.

Aurélien Garivier and Emilie Kaufmann. Optimal best arm identification with fixed confidence. In *Conference on Learning Theory*, pages 998–1027. PMLR, 2016.

Joy He-Yueya, Jonathan Lee, Matthew Jörke, and Emma Brunskill. Cost-aware near-optimal policy learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 28088–28096, 2025.

Xiaoyan Hu, Lauren Pick, Ho-fung Leung, and Farzan Farnia. Promptwise: Online learning for cost-aware prompt assignment in generative models. *arXiv preprint arXiv:2505.18901*, 2025.

Alexander D Ioffe. Variational analysis of regular mappings. *Springer Monographs in Mathematics. Springer, Cham*, 2017.

Marc Jourdan. *Solving pure exploration problems with the Top Two approach*. PhD thesis, Université de Lille, 2024.

Marc Jourdan, Rémy Degenne, Dorian Baudry, Rianne de Heide, and Emilie Kaufmann. Top two algorithms revisited. *Advances in Neural Information Processing Systems*, 35:26791–26803, 2022.

Kellen Kanarios, Qining Zhang, and Lei Ying. Cost aware best arm identification. *Reinforcement Learning Journal*, 2024.

Andrea Locatelli, Maurilio Gutzeit, and Alexandra Carpentier. An optimal algorithm for the thresholding bandit problem. In *International Conference on Machine Learning*, pages 1690–1698. PMLR, 2016.

Pierre Ménard. Gradient ascent for active exploration in bandit problems. *arXiv preprint arXiv:1905.08165*, 2019.

Paul Milgrom and Ilya Segal. Envelope theorems for arbitrary choice sets. *Econometrica*, 70(2):583–601, 2002.

Riccardo Poiani, Rémy Degenne, Emilie Kaufmann, Alberto Maria Metelli, and Marcello Restelli. Optimal multi-fidelity best-arm identification. *Advances in Neural Information Processing Systems*, 37:121882–121927, 2024.

Chao Qin and Daniel Russo. Optimizing adaptive experiments: A unified approach to regret minimization and best-arm identification. *arXiv preprint arXiv:2402.10592*, 2024.

Chao Qin and Wei You. Dual-directed algorithm design for efficient pure exploration. *Operations Research*, 2025.

Chao Qin, Diego Klabjan, and Daniel Russo. Improving the expected improvement algorithm. *Advances in Neural Information Processing Systems*, 30, 2017.

Daniel Russo. Simple bayesian algorithms for best-arm identification. *Operations Research*, 68(6):1625–1647, 2020.

Xuedong Shang, Rianne Heide, Pierre Ménard, Emilie Kaufmann, and Michal Valko. Fixed-confidence guarantees for Bayesian best-arm identification. In *International Conference on Artificial Intelligence and Statistics*, pages 1823–1832. PMLR, 2020.

Marta Soare, Alessandro Lazaric, and Rémi Munos. Best-arm identification in linear bandits. *Advances in neural information processing systems*, 27, 2014.

Po-An Wang, Ruo-Chun Tzeng, and Alexandre Proutiere. Fast pure exploration via Frank-Wolfe. *Advances in Neural Information Processing Systems*, 34:5810–5821, 2021.

Di Wu, Chengshuai Shi, Ruida Zhou, and Cong Shen. Cost-aware optimal pairwise pure exploration. *arXiv preprint arXiv:2503.07877*, 2025.

Liyuan Xu, Junya Honda, and Masashi Sugiyama. A fully adaptive algorithm for pure exploration in linear bandits. In *International Conference on Artificial Intelligence and Statistics*, pages 843–851. PMLR, 2018.

Junwen Yang and Vincent Tan. Minimax optimal fixed-budget best arm identification in linear bandits. *Advances in Neural Information Processing Systems*, 35:12253–12266, 2022.

Wei You, Chao Qin, Zihao Wang, and Shuoguang Yang. Information-directed selection for top-two algorithms. In *The Thirty Sixth Annual Conference on Learning Theory*, pages 2850–2851. PMLR, 2023.

Appendix

A Proof of Theorem 1

Our proof follows the basic template of Russo (2020, Theorem 1), adapted to the cost-aware and structural bandit setting. The execution is relatively routine, nevertheless, we include full arguments for completeness.

A.1 Preliminaries

Recall the average KL divergence under a probability vector $\boldsymbol{p} \in \Delta_K$:

$$\Gamma(\boldsymbol{p}; \boldsymbol{\vartheta}) \triangleq \sum_{i=1}^{K} p_i \mathrm{KL}(P_{\boldsymbol{\theta},i} \| P_{\boldsymbol{\vartheta},i}).$$

Let the log-likelihood ratio up to time n be

$$\Lambda_n(\boldsymbol{\theta} \| \boldsymbol{\vartheta}) \triangleq \sum_{\ell=1}^{n} \log \frac{p_{\boldsymbol{\theta},I_\ell}(Y_\ell)}{p_{\boldsymbol{\vartheta},I_\ell}(Y_\ell)}.$$

Lemma 3 (Uniform log-likelihood approximation)**.** *Under Assumption 1,*

$$\lim_{n\to\infty} \sup_{\boldsymbol{\vartheta}\in\Theta} \left| \frac{1}{n} \Lambda_n(\boldsymbol{\theta} \| \boldsymbol{\vartheta}) - \Gamma(\bar{\boldsymbol{\psi}}_n; \boldsymbol{\vartheta}) \right| = 0 \qquad \textit{a.s.}$$

Proof. Fix the true instance $\boldsymbol{\theta}$. For each fixed $\boldsymbol{\vartheta} \in \Theta$, define

$$X_\ell(\boldsymbol{\vartheta}) \triangleq \log \frac{p_{\boldsymbol{\theta},I_\ell}(Y_\ell)}{p_{\boldsymbol{\vartheta},I_\ell}(Y_\ell)}.$$

Then $X_\ell(\boldsymbol{\vartheta})$ is $\mathcal{H}_\ell$-measurable and

$$\mathbb{E}_{\boldsymbol{\theta}}[X_\ell(\boldsymbol{\vartheta}) \mid \mathcal{H}_{\ell-1}] = \sum_{i=1}^{K} \psi_{\ell,i} \mathbb{E}_{\boldsymbol{\theta}} \left[\log \frac{p_{\boldsymbol{\theta},i}(Y)}{p_{\boldsymbol{\vartheta},i}(Y)} \right] = \sum_{i=1}^{K} \psi_{\ell,i} \mathrm{KL}(P_{\boldsymbol{\theta},i} \| P_{\boldsymbol{\vartheta},i}).$$

Therefore,

$$\Lambda_n(\boldsymbol{\theta} \| \boldsymbol{\vartheta}) = \sum_{\ell=1}^{n} X_\ell(\boldsymbol{\vartheta}) = \sum_{\ell=1}^{n} \mathbb{E}_{\boldsymbol{\theta}}[X_\ell(\boldsymbol{\vartheta}) \mid \mathcal{H}_{\ell-1}] + M_n(\boldsymbol{\vartheta}) = n\Gamma(\bar{\boldsymbol{\psi}}_n; \boldsymbol{\vartheta}) + M_n(\boldsymbol{\vartheta}),$$

where $M_n(\boldsymbol{\vartheta})$ is the martingale

$$M_n(\boldsymbol{\vartheta}) \triangleq \sum_{\ell=1}^{n} \delta_\ell(\boldsymbol{\vartheta}), \qquad \delta_\ell(\boldsymbol{\vartheta}) \triangleq X_\ell(\boldsymbol{\vartheta}) - \mathbb{E}_{\boldsymbol{\theta}}\left[X_\ell(\boldsymbol{\vartheta}) \mid \mathcal{H}_{\ell-1}\right].$$

Thus it remains to show that $\sup_{\boldsymbol{\vartheta}\in\Theta} |M_n(\boldsymbol{\vartheta})| = o(n)$ a.s.

Under Assumption 1, we have a finite uniform bound

$$\kappa_{\mathrm{KL}} \triangleq \sup_{i\in[K]} \sup_{\boldsymbol{\vartheta}\in\Theta} \mathrm{KL}(P_{\boldsymbol{\theta},i} \| P_{\boldsymbol{\vartheta},i}) \le \sup_{i\in[K]} \sup_{\boldsymbol{\vartheta}\in\Theta} \left(\mathbb{E}_{\boldsymbol{\theta}} \left[\left(\log \frac{p_{\boldsymbol{\theta},i}(Y)}{p_{\boldsymbol{\vartheta},i}(Y)} \right)^2 \right] \right)^{1/2} < \infty. \tag{19}$$

In particular, for each fixed $\boldsymbol{\vartheta}$, $\{M_n(\boldsymbol{\vartheta})\}_{n\geq 1}$ is square-integrable and

$$\mathbb{E}_{\boldsymbol{\theta}}\left[\delta_\ell(\boldsymbol{\vartheta})^2 \mid \mathcal{H}_{\ell-1}\right] \leq 4\mathbb{E}_{\boldsymbol{\theta}}\left[X_\ell(\boldsymbol{\vartheta})^2 \mid \mathcal{H}_{\ell-1}\right] \leq 4\kappa_{\mathrm{KL}}^2,$$

where we used $(a - \mathbb{E}[a \mid \mathcal{H}])^2 \leq 4a^2$ and the definition of κ_X. Hence,

$$\sum_{\ell=1}^{\infty} \frac{\mathbb{E}_{\boldsymbol{\theta}}[\delta_\ell(\boldsymbol{\vartheta})^2]}{\ell^2} \leq 4\kappa_{\mathrm{KL}}^2 \sum_{\ell=1}^{\infty} \frac{1}{\ell^2} < \infty.$$

By the strong law of large numbers for square-integrable martingales, this implies $M_n(\boldsymbol{\vartheta})/n \to 0$ a.s. for each fixed $\boldsymbol{\vartheta} \in \Theta$. Next fix $\delta > 0$. By Assumption 1, there exists a finite δ-net $\mathcal{N}_\delta \subset \Theta$. Since $\mathcal{N}_\delta$ is finite and $M_n(\boldsymbol{\nu})/n \to 0$ a.s. for each $\boldsymbol{\nu} \in \mathcal{N}_\delta$, we have

$$\max_{\boldsymbol{\nu}\in\mathcal{N}_\delta} \frac{|M_n(\boldsymbol{\nu})|}{n} \to 0 \qquad \text{a.s.} \tag{20}$$

We now control $|M_n(\boldsymbol{\vartheta}) - M_n(\boldsymbol{\vartheta}')|$ for nearby parameters. Assumption 1 implies the following Lipschitz property: there exists a finite constant $L < \infty$ such that for all $i \in [K]$, all $\boldsymbol{\vartheta}, \boldsymbol{\vartheta}' \in \Theta$, and ν-a.e. y,

$$\left|\log \frac{p_{\boldsymbol{\vartheta}',i}(y)}{p_{\boldsymbol{\vartheta},i}(y)}\right| \leq L\|\boldsymbol{\vartheta} - \boldsymbol{\vartheta}'\|(1 + |T(y)|). \tag{21}$$

Using (21) with $i = I_\ell$ and $y = Y_\ell$ yields, for all $\boldsymbol{\vartheta}, \boldsymbol{\vartheta}' \in \Theta$,

$$|X_\ell(\boldsymbol{\vartheta}) - X_\ell(\boldsymbol{\vartheta}')| = \left|\log \frac{p_{\boldsymbol{\vartheta}',I_\ell}(Y_\ell)}{p_{\boldsymbol{\vartheta},I_\ell}(Y_\ell)}\right| \leq L\|\boldsymbol{\vartheta} - \boldsymbol{\vartheta}'\|(1 + |T(Y_\ell)|). \tag{22}$$

By Jensen's inequality,

$$\begin{aligned}\left|\mathbb{E}_{\boldsymbol{\theta}}\left[X_\ell(\boldsymbol{\vartheta}) - X_\ell(\boldsymbol{\vartheta}') \mid \mathcal{H}_{\ell-1}\right]\right| &\leq \mathbb{E}_{\boldsymbol{\theta}}\left[|X_\ell(\boldsymbol{\vartheta}) - X_\ell(\boldsymbol{\vartheta}')| \mid \mathcal{H}_{\ell-1}\right] \\ &\leq L\|\boldsymbol{\vartheta} - \boldsymbol{\vartheta}'\|\Big(1 + \mathbb{E}_{\boldsymbol{\theta}}[|T(Y_\ell)| \mid \mathcal{H}_{\ell-1}]\Big).\end{aligned} \tag{23}$$

Combining (22)–(23) and the definition of $\delta_\ell(\cdot)$ gives

$$|\delta_\ell(\boldsymbol{\vartheta}) - \delta_\ell(\boldsymbol{\vartheta}')| \leq |X_\ell(\boldsymbol{\vartheta}) - X_\ell(\boldsymbol{\vartheta}')| + \left|\mathbb{E}_{\boldsymbol{\theta}}[X_\ell(\boldsymbol{\vartheta}) - X_\ell(\boldsymbol{\vartheta}') \mid \mathcal{H}_{\ell-1}]\right|$$

and hence

$$|\delta_\ell(\boldsymbol{\vartheta}) - \delta_\ell(\boldsymbol{\vartheta}')| \leq L\|\boldsymbol{\vartheta} - \boldsymbol{\vartheta}'\|\Big(2 + |T(Y_\ell)| + \mathbb{E}_{\boldsymbol{\theta}}[|T(Y_\ell)| \mid \mathcal{H}_{\ell-1}]\Big). \tag{24}$$

Summing (24) over $\ell \leq n$ yields

$$\frac{|M_n(\boldsymbol{\vartheta}) - M_n(\boldsymbol{\vartheta}')|}{n} \leq L\|\boldsymbol{\vartheta} - \boldsymbol{\vartheta}'\|\left(2 + \frac{1}{n}\sum_{\ell=1}^{n}|T(Y_\ell)| + \frac{1}{n}\sum_{\ell=1}^{n}\mathbb{E}_{\boldsymbol{\theta}}[|T(Y_\ell)| \mid \mathcal{H}_{\ell-1}]\right). \tag{25}$$

Define the constants

$$\kappa_T^{(2)} \triangleq \sup_{i\in[K]} \mathbb{E}_{\boldsymbol{\theta}}\left[T(Y)^2\right] < \infty, \qquad \kappa_T^{(1)} \triangleq \sup_{i\in[K]} \mathbb{E}_{\boldsymbol{\theta}}\left[|T(Y)|\right] \leq \sqrt{\kappa_T^{(2)}} < \infty.$$

Then for every ℓ,

$$\mathbb{E}_{\boldsymbol{\theta}}[|T(Y_\ell)| \mid \mathcal{H}_{\ell-1}] = \sum_{i=1}^{K} \psi_{\ell,i} \mathbb{E}_{\boldsymbol{\theta}}[|T(Y)| \mid I_\ell = i] \le \kappa_T^{(1)},$$

and therefore

$$\frac{1}{n}\sum_{\ell=1}^{n} \mathbb{E}_{\boldsymbol{\theta}}[|T(Y_\ell)| \mid \mathcal{H}_{\ell-1}] \le \kappa_T^{(1)}. \tag{26}$$

Moreover, letting

$$U_\ell \triangleq |T(Y_\ell)| - \mathbb{E}_{\boldsymbol{\theta}}[|T(Y_\ell)| \mid \mathcal{H}_{\ell-1}], \qquad S_n \triangleq \frac{1}{n}\sum_{\ell=1}^{n} U_\ell,$$

we have that $\{U_\ell\}_{\ell \ge 1}$ is a martingale difference sequence and

$$\mathbb{E}_{\boldsymbol{\theta}}[U_\ell^2 \mid \mathcal{H}_{\ell-1}] \le \mathbb{E}_{\boldsymbol{\theta}}[|T(Y_\ell)|^2 \mid \mathcal{H}_{\ell-1}] = \sum_{i=1}^{K} \psi_{\ell,i} \mathbb{E}_{\boldsymbol{\theta}}[T(Y)^2 \mid I_\ell = i] \le \kappa_T^{(2)}.$$

Hence $\sum_{\ell \ge 1} \mathbb{E}_{\boldsymbol{\theta}}[U_\ell^2]/\ell^2 < \infty$, and by the martingale strong law,

$$S_n \to 0 \qquad \text{a.s.} \tag{27}$$

Finally,

$$\frac{1}{n}\sum_{\ell=1}^{n} |T(Y_\ell)| = \frac{1}{n}\sum_{\ell=1}^{n} \mathbb{E}_{\boldsymbol{\theta}}[|T(Y_\ell)| \mid \mathcal{H}_{\ell-1}] + S_n \le \kappa_T^{(1)} + S_n. \tag{28}$$

Plugging (26) and (28) into (25) gives

$$\frac{|M_n(\boldsymbol{\vartheta}) - M_n(\boldsymbol{\vartheta}')|}{n} \le L\|\boldsymbol{\vartheta} - \boldsymbol{\vartheta}'\|(2 + 2\kappa_T^{(1)} + S_n) \qquad \text{a.s.} \tag{29}$$

Fix $\delta > 0$ and let $\mathcal{N}_\delta$ be a finite δ-net of Θ. For each $\boldsymbol{\vartheta} \in \Theta$, choose $\pi_\delta(\boldsymbol{\vartheta}) \in \mathcal{N}_\delta$ such that $\|\boldsymbol{\vartheta} - \pi_\delta(\boldsymbol{\vartheta})\| \le \delta$. Then

$$\frac{|M_n(\boldsymbol{\vartheta})|}{n} \le \frac{|M_n(\pi_\delta(\boldsymbol{\vartheta}))|}{n} + \frac{|M_n(\boldsymbol{\vartheta}) - M_n(\pi_\delta(\boldsymbol{\vartheta}))|}{n}.$$

Taking $\sup_{\boldsymbol{\vartheta} \in \Theta}$ and using (29) yields

$$\begin{aligned}
\sup_{\boldsymbol{\vartheta}\in\Theta} \frac{|M_n(\boldsymbol{\vartheta})|}{n} &\le \max_{\boldsymbol{\nu}\in\mathcal{N}_\delta} \frac{|M_n(\boldsymbol{\nu})|}{n} + \sup_{\substack{\boldsymbol{\vartheta}\in\Theta: \\ \|\boldsymbol{\vartheta}-\pi_\delta(\boldsymbol{\vartheta})\|\le\delta}} L\|\boldsymbol{\vartheta} - \pi_\delta(\boldsymbol{\vartheta})\|(2 + 2\kappa_T^{(1)} + |S_n|) \\
&\le \max_{\boldsymbol{\nu}\in\mathcal{N}_\delta} \frac{|M_n(\boldsymbol{\nu})|}{n} + L\delta(2 + 2\kappa_T^{(1)}) + L\delta|S_n|.
\end{aligned} \tag{30}$$

Now let $n \to \infty$ in (30). Using (20) and $S_n \to 0$ a.s. yields

$$\limsup_{n\to\infty} \sup_{\boldsymbol{\vartheta}\in\Theta} \frac{|M_n(\boldsymbol{\vartheta})|}{n} \le L\delta(2 + 2\kappa_T^{(1)}) \qquad \text{a.s.}$$

Since $\delta > 0$ is arbitrary, letting $\delta \to 0$ gives $\sup_{\boldsymbol{\vartheta}\in\Theta} |M_n(\boldsymbol{\vartheta})|/n \to 0$ a.s. Therefore,

$$\sup_{\boldsymbol{\vartheta}\in\Theta} \left| \frac{1}{n}\Lambda_n(\boldsymbol{\theta}\|\boldsymbol{\vartheta}) - \Gamma(\bar{\boldsymbol{\psi}}_n; \boldsymbol{\vartheta}) \right| = \sup_{\boldsymbol{\vartheta}\in\Theta} \frac{|M_n(\boldsymbol{\vartheta})|}{n} \to 0 \qquad \text{a.s.},$$

which proves the lemma. □

Lemma 4 (Uniform posterior density ratio)**.** *Assume Assumptions 1–4. Then,*

$$\sup_{\boldsymbol{\vartheta}\in\Theta}\left|\frac{1}{n}\log\frac{\pi_n(\boldsymbol{\vartheta})}{\pi_n(\boldsymbol{\theta})}+\Gamma(\bar{\boldsymbol{\psi}}_n;\boldsymbol{\vartheta})\right|\to 0 \qquad a.s.$$

Proof. By Bayes' rule,

$$\log\frac{\pi_n(\boldsymbol{\vartheta})}{\pi_n(\boldsymbol{\theta})}=\log\frac{\pi_0(\boldsymbol{\vartheta})}{\pi_0(\boldsymbol{\theta})}+\sum_{\ell=1}^{n}\log\frac{p_{\boldsymbol{\vartheta},I_\ell}(Y_\ell)}{p_{\boldsymbol{\theta},I_\ell}(Y_\ell)}=\log\frac{\pi_0(\boldsymbol{\vartheta})}{\pi_0(\boldsymbol{\theta})}-\Lambda_n(\boldsymbol{\theta}\|\boldsymbol{\vartheta}).$$

Divide by n, we obtain

$$\frac{1}{n}\log\frac{\pi_n(\boldsymbol{\vartheta})}{\pi_n(\boldsymbol{\theta})}+\Gamma(\bar{\boldsymbol{\psi}}_n;\boldsymbol{\vartheta})=\frac{1}{n}\log\frac{\pi_0(\boldsymbol{\vartheta})}{\pi_0(\boldsymbol{\theta})}-\left(\frac{1}{n}\Lambda_n(\boldsymbol{\theta}\|\boldsymbol{\vartheta})-\Gamma(\bar{\boldsymbol{\psi}}_n;\boldsymbol{\vartheta})\right).$$

The first term goes to 0 uniformly in $\boldsymbol{\vartheta}$ by Assumption 4 and the second term goes to 0 uniformly by Lemma 3. □

Next we need a mild regularity of $\boldsymbol{\vartheta}\mapsto\Gamma(\bar{\boldsymbol{\psi}}_n;\boldsymbol{\vartheta})$.

Lemma 5 (Uniform continuity of $\Gamma(\boldsymbol{q};\boldsymbol{\vartheta})$ in $\boldsymbol{\vartheta}$)**.** *Assume Assumptions 1 and 1. Then for every $\epsilon>0$ there exists $\delta>0$ such that for all $\boldsymbol{q}\in\Delta_K$ and all $\boldsymbol{\vartheta},\boldsymbol{\vartheta}'\in\Theta$,*

$$\|\boldsymbol{\vartheta}-\boldsymbol{\vartheta}'\|\le\delta \quad\Rightarrow\quad |\Gamma(\boldsymbol{q};\boldsymbol{\vartheta})-\Gamma(\boldsymbol{q};\boldsymbol{\vartheta})|\le\epsilon.$$

Proof. For each arm i, continuity of η_i on compact Θ implies uniform continuity. Recall that

$$\mathrm{KL}(P_{\boldsymbol{\theta},i}\|P_{\boldsymbol{\vartheta},i})=(\eta_i(\boldsymbol{\theta})-\eta_i(\boldsymbol{\vartheta}))A'(\eta_i(\boldsymbol{\theta}))-A(\eta_i(\boldsymbol{\theta}))+A(\eta_i(\boldsymbol{\vartheta})).$$

On compact $\eta_i(\Theta)$, A is Lipschitz as A' is bounded by Assumption 1. Thus $\boldsymbol{\vartheta}\mapsto\mathrm{KL}(P_{\boldsymbol{\theta},i}\|P_{\boldsymbol{\vartheta},i})$ is uniformly continuous, uniformly in i. A convex combination over i with weights q_i preserves uniform continuity uniformly over $q\in\Delta_K$. □

A.2 Posterior Large Deviation

Definition 1 (Logarithmic equivalence)**.** *For positive sequences $\{a_n\}$ and $\{b_n\}$, write $a_n\doteq b_n$ if $\lim_{n\to\infty}\frac{1}{n}\log(a_n/b_n)=0$. If the sequences are random, the relation holds almost surely.*

Proposition 2 (Posterior large deviations)**.** *Assume Assumptions 1–4. Let $\tilde{\Theta}\subset\Theta$ be open. Then,*

$$\Pi_n(\tilde{\Theta})\ \doteq\ \exp\Big(-n\inf_{\boldsymbol{\vartheta}\in\tilde{\Theta}}\Gamma(\bar{\boldsymbol{\psi}}_n;\boldsymbol{\vartheta})\Big). \tag{31}$$

Proof. Fix $\tilde{\Theta}$ open. Write

$$\Pi_n(\tilde{\Theta})=\frac{\int_{\tilde{\Theta}}\pi_n(\boldsymbol{\vartheta})\,\mathrm{d}\boldsymbol{\vartheta}}{\int_{\Theta}\pi_n(\boldsymbol{\vartheta})\,\mathrm{d}\boldsymbol{\vartheta}}=\frac{\int_{\tilde{\Theta}}\frac{\pi_n(\boldsymbol{\vartheta})}{\pi_n(\boldsymbol{\theta})}\,\mathrm{d}\boldsymbol{\vartheta}}{\int_{\Theta}\frac{\pi_n(\boldsymbol{\vartheta})}{\pi_n(\boldsymbol{\theta})}\,\mathrm{d}\boldsymbol{\vartheta}}.$$

By Lemma 4, for some deterministic $\epsilon_n\downarrow 0$,

$$\exp\{-n(\Gamma(\bar{\boldsymbol{\psi}}_n;\boldsymbol{\vartheta})+\epsilon_n)\}\le\frac{\pi_n(\boldsymbol{\vartheta})}{\pi_n(\boldsymbol{\theta})}\le\exp\{-n(\Gamma(\bar{\boldsymbol{\psi}}_n;\boldsymbol{\vartheta})-\epsilon_n)\}\qquad\forall\boldsymbol{\vartheta}\in\Theta.$$

Integrating over $\tilde{\Theta}$ and Θ yields that the numerator and denominator are each log-equivalent to the corresponding Laplace integrals

$$\int_{\tilde{\Theta}} \exp\{-n\Gamma(\bar{\boldsymbol{\psi}}_n; \boldsymbol{\vartheta})\} \, \mathrm{d}\boldsymbol{\vartheta} \quad \text{and} \quad \int_{\Theta} \exp\{-n\Gamma(\bar{\boldsymbol{\psi}}_n; \boldsymbol{\vartheta})\} \, \mathrm{d}\boldsymbol{\vartheta}.$$

Thus it suffices to show the Laplace principle

$$\int_{\tilde{\Theta}} \exp\{-nW_n(\boldsymbol{\vartheta})\} \, \mathrm{d}\boldsymbol{\vartheta} \doteq \exp\{-n \inf_{\boldsymbol{\vartheta} \in \tilde{\Theta}} W_n(\boldsymbol{\vartheta})\} \quad \text{with} \quad W_n(\boldsymbol{\vartheta}) = \Gamma(\bar{\boldsymbol{\psi}}_n; \boldsymbol{\vartheta}),$$

and similarly with $\tilde{\Theta} = \Theta$.

Let $\hat{\boldsymbol{\vartheta}}_n \in \mathrm{cl}(\tilde{\Theta})$ attain the minimum $W_n(\hat{\boldsymbol{\vartheta}}_n) = \inf_{\boldsymbol{\vartheta} \in \tilde{\Theta}} W_n(\boldsymbol{\vartheta})$, which exists because W_n is continuous (Lemma 5) and $\mathrm{cl}(\tilde{\Theta})$ is compact. Define $\gamma_n = \int_{\tilde{\Theta}} \exp(-nW_n(\boldsymbol{\vartheta})) \, \mathrm{d}\boldsymbol{\vartheta}$. Then

$$\gamma_n \leq \mathrm{Vol}(\tilde{\Theta}) \exp\{-nW_n(\hat{\boldsymbol{\vartheta}}_n)\},$$

so $\limsup_{n\to\infty} \frac{1}{n} \log \gamma_n + W_n(\hat{\boldsymbol{\vartheta}}_n) \leq 0$.

For the reverse bound, fix $\epsilon > 0$. By uniform continuity of W_n (Lemma 5), there exists $\delta > 0$ such that $\|\boldsymbol{\vartheta} - \boldsymbol{\vartheta}'\| \leq \delta$ implies $|W_n(\boldsymbol{\vartheta}) - W_n(\boldsymbol{\vartheta}')| \leq \epsilon$ for all n. Because $\tilde{\Theta}$ is open, for each n we can choose a point $\boldsymbol{\vartheta}_n^\circ \in \tilde{\Theta}$ arbitrarily close to $\hat{\boldsymbol{\vartheta}}_n$. Then the ball $B(\boldsymbol{\vartheta}_n^\circ, \delta/2)$ has positive intersection with $\tilde{\Theta}$ and volume bounded below by some $c_\delta > 0$ uniformly in n (compactness plus finiteness of a cover argument). On that intersection, $W_n(\boldsymbol{\vartheta}) \leq W_n(\hat{\boldsymbol{\vartheta}}_n) + \epsilon$. Therefore

$$\gamma_n \geq \int_{\tilde{\Theta} \cap B(\boldsymbol{\vartheta}_n^\circ, \delta/2)} \exp\{-nW_n(\boldsymbol{\vartheta})\} \, \mathrm{d}\boldsymbol{\vartheta} \geq c_\delta \exp\{-n(W_n(\hat{\boldsymbol{\vartheta}}_n) + \epsilon)\}.$$

Taking logs and dividing by n gives

$$\frac{1}{n} \log \gamma_n + W_n(\hat{\boldsymbol{\vartheta}}_n) \geq \frac{1}{n} \log c_\delta - \epsilon \to -\epsilon.$$

Since ϵ is arbitrary, $\liminf_{n\to\infty} \frac{1}{n} \log \gamma_n + W_n(\hat{\boldsymbol{\vartheta}}_n) \geq 0$. This establishes the Laplace principle for $\tilde{\Theta}$.

For Θ itself, note that $\inf_{\boldsymbol{\vartheta} \in \Theta} W_n(\boldsymbol{\vartheta}) = 0$ because $W_n(\boldsymbol{\theta}) = 0$ and $W_n \geq 0$. Thus the denominator integral is log-equivalent to $\exp(0) = 1$ and cancels. This yields (31). □

A.3 From Sample-Normalized to Cost-Normalized Exponents

Lemma 6. *For each arm i, we have $\lim_{n\to\infty}(p_{n,i} - \bar{\psi}_{n,i}) = 0$ a.s. Consequently, $\|\boldsymbol{p}_n - \bar{\boldsymbol{\psi}}_n\|_1 \to 0$ a.s.*

Proof. Let $X_\ell = \mathbb{1}\{I_\ell = i\}$ and $Z_\ell = \mathbb{E}[X_\ell \mid \mathcal{H}_{\ell-1}] = \psi_{\ell,i}$. Then $M_n = \sum_{\ell=1}^n (X_\ell - Z_\ell)$ is a martingale with bounded increments, hence $M_n/n \to 0$ a.s. □

Proposition 3 (Cost-normalized posterior exponent)**.** *Assume Assumptions 1–4. Then for any open $\tilde{\Theta} \subset \Theta$,*

$$-\frac{1}{B_n} \log \Pi_n(\tilde{\Theta}) \;\doteq\; \inf_{\boldsymbol{\vartheta} \in \tilde{\Theta}} \frac{\Gamma(\boldsymbol{p}_n; \boldsymbol{\vartheta})}{\bar{C}(\boldsymbol{p}_n)} \qquad \text{a.s.}$$

Proof. From Proposition 2,

$$-\frac{1}{n}\log \Pi_n(\tilde{\Theta}) \;\doteq\; \inf_{\boldsymbol{\vartheta}\in\tilde{\Theta}} \Gamma(\bar{\boldsymbol{\psi}}_n;\boldsymbol{\vartheta}). \tag{32}$$

For any $\boldsymbol{\vartheta} \in \Theta$,

$$\Gamma(\boldsymbol{p}_n;\boldsymbol{\vartheta}) - \Gamma(\bar{\boldsymbol{\psi}}_n;\boldsymbol{\vartheta}) = \sum_{i=1}^{K}(p_{n,i} - \bar{\psi}_{n,i})\mathrm{KL}(P_{\boldsymbol{\theta},i}\|P_{\boldsymbol{\vartheta},i}),$$

hence by (19),

$$\sup_{\boldsymbol{\vartheta}\in\Theta}\left|\Gamma(\boldsymbol{p}_n;\boldsymbol{\vartheta}) - \Gamma(\bar{\boldsymbol{\psi}}_n;\boldsymbol{\vartheta})\right| \le \kappa_{\mathrm{KL}}\|\boldsymbol{p}_n - \bar{\boldsymbol{\psi}}_n\|_1 \longrightarrow 0 \qquad \text{a.s.},$$

where the convergence is Lemma 6. Therefore,

$$\left|\inf_{\boldsymbol{\vartheta}\in\tilde{\Theta}} \Gamma(\boldsymbol{p}_n;\boldsymbol{\vartheta}) - \inf_{\boldsymbol{\vartheta}\in\tilde{\Theta}} \Gamma(\bar{\boldsymbol{\psi}}_n;\boldsymbol{\vartheta})\right| \le \sup_{\boldsymbol{\vartheta}\in\Theta}\left|\Gamma(\boldsymbol{p}_n;\boldsymbol{\vartheta}) - \Gamma(\bar{\boldsymbol{\psi}}_n;\boldsymbol{\vartheta})\right| \longrightarrow 0 \qquad \text{a.s.} \tag{33}$$

Combining (32) and (33) yields

$$\Pi_n(\tilde{\Theta}) \;\doteq\; \exp\left\{-n \inf_{\boldsymbol{\vartheta}\in\tilde{\Theta}} \Gamma(\boldsymbol{p}_n;\boldsymbol{\vartheta})\right\} \qquad \text{a.s.},$$

Multiply by $n/B_n = 1/\bar{C}(\boldsymbol{p}_n)$ to obtain

$$-\frac{1}{B_n}\log \Pi_n(\tilde{\Theta}) \;\doteq\; \inf_{\boldsymbol{\vartheta}\in\tilde{\Theta}} \frac{\Gamma(\boldsymbol{p}_n;\boldsymbol{\vartheta})}{\bar{C}(\boldsymbol{p}_n)},$$

using that multiplying by deterministic sequences bounded away from 0 and ∞ preserves log-equivalence. The second statement follows if the right-hand side converges. □

A.4 Completing the Proof

We are now ready to prove Theorem 1.

Proof. We prove the upper bound first. Fix $\epsilon > 0$ and let $\tilde{\Theta}_\epsilon$ be the open superset of $\mathrm{Alt}(\boldsymbol{\theta})$ from Assumption 3. Then $\Pi_n(\mathrm{Alt}(\boldsymbol{\theta})) \le \Pi_n(\tilde{\Theta}_\epsilon)$, hence

$$-\frac{1}{B_n}\log \Pi_n(\mathrm{Alt}(\boldsymbol{\theta})) \;\ge\; -\frac{1}{B_n}\log \Pi_n(\tilde{\Theta}_\epsilon).$$

This direction is not useful for an upper bound; instead we use $\Pi_n(\mathrm{Alt}(\boldsymbol{\theta})) \ge \Pi_n(\tilde{\Theta})$ for any open $\tilde{\Theta} \subset \mathrm{Alt}(\boldsymbol{\theta})$. Let $\tilde{\Theta}$ be any open subset of $\mathrm{Alt}(\boldsymbol{\theta})$ (e.g. an interior approximation). Then Proposition 3 yields

$$-\frac{1}{B_n}\log \Pi_n(\mathrm{Alt}(\boldsymbol{\theta})) \le -\frac{1}{B_n}\log \Pi_n(\tilde{\Theta}) \doteq \inf_{\boldsymbol{\vartheta}\in\tilde{\Theta}} \frac{\Gamma(\boldsymbol{p}_n;\boldsymbol{\vartheta})}{\bar{C}(\boldsymbol{p}_n)}.$$

Taking lim sup in n and using $\tilde{\Theta} \subset \mathrm{Alt}(\boldsymbol{\theta})$ gives

$$\limsup_{n\to\infty} -\frac{1}{B_n}\log \Pi_n(\mathrm{Alt}(\boldsymbol{\theta})) \le \limsup_{n\to\infty} \inf_{\boldsymbol{\vartheta}\in\mathrm{Alt}(\boldsymbol{\theta})} \frac{\Gamma(\boldsymbol{p}_n;\boldsymbol{\vartheta})}{\bar{C}(\boldsymbol{p}_n)}.$$

Finally, since for each n, $\boldsymbol{p}_n \in \Delta_K$,

$$\inf_{\boldsymbol{\vartheta}\in \mathrm{Alt}(\boldsymbol{\theta})} \frac{\Gamma(\boldsymbol{p}_n;\boldsymbol{\vartheta})}{\bar{C}(\boldsymbol{p}_n)} \le \sup_{\boldsymbol{p}\in\Delta_K} \inf_{\boldsymbol{\vartheta}\in \mathrm{Alt}(\boldsymbol{\theta})} \frac{\Gamma(\boldsymbol{p};\boldsymbol{\vartheta})}{\bar{C}(\boldsymbol{p})} = \Gamma^*.$$

Thus the lim sup is bounded by Γ^*.

For sufficiency, recall that

$$\inf_{\boldsymbol{\vartheta}\in \mathrm{Alt}(\boldsymbol{\theta})} \frac{\Gamma(\boldsymbol{p}_n;\boldsymbol{\vartheta})}{\bar{C}(\boldsymbol{p}_n)} = \inf_{\boldsymbol{\vartheta}\in \mathrm{Alt}(\boldsymbol{\theta})} \Gamma^{\mathrm{c}}(\boldsymbol{p}_n;\boldsymbol{\vartheta}),$$

Assume (3) holds, then so the right-hand side above converges to Γ^* a.s. Now apply Proposition 3 to the (open) sets $\tilde{\Theta}_\epsilon$ that approximate $\mathrm{Alt}(\boldsymbol{\theta})$. Because $\mathrm{Alt}(\boldsymbol{\theta}) \subset \tilde{\Theta}_\epsilon$,

$$-\frac{1}{B_n}\log \Pi_n(\mathrm{Alt}(\boldsymbol{\theta})) \ge -\frac{1}{B_n}\log \Pi_n(\tilde{\Theta}_\epsilon) \doteq \inf_{\boldsymbol{\vartheta}\in\tilde{\Theta}_\epsilon} \frac{\Gamma(\bar{\boldsymbol{\psi}}_n;\boldsymbol{\vartheta})}{\bar{C}(\boldsymbol{p}_n)}.$$

Taking lim inf in n and then letting $\epsilon \downarrow 0$, we obtain

$$\liminf_{n\to\infty} -\frac{1}{B_n}\log \Pi_n(\mathrm{Alt}(\boldsymbol{\theta})) \ge \lim_{\epsilon\downarrow 0} \liminf_{n\to\infty} \inf_{\boldsymbol{\vartheta}\in\tilde{\Theta}_\epsilon} \frac{\Gamma(\bar{\boldsymbol{\psi}}_n;\boldsymbol{\vartheta})}{\bar{C}(\boldsymbol{p}_n)} = \Gamma^*,$$

where the last equality uses that $\tilde{\Theta}_\epsilon$ shrinks to $\mathrm{Alt}(\boldsymbol{\theta})$ and the objective is continuous in $\boldsymbol{\vartheta}$ by Lemma 5 and in $\boldsymbol{p}$; thus the infimum over $\tilde{\Theta}_\epsilon$ converges to the infimum over $\mathrm{Alt}(\boldsymbol{\theta})$.

Combine this lower bound with the upper bound conclude that the limit exists and equals Γ^*. □

B Proof of Lemma 1

Fix $\boldsymbol{\theta}$ and $x \in \mathcal{X}(\boldsymbol{\theta})$. Recall

$$\tilde{\Gamma}^c(\boldsymbol{w},\boldsymbol{\vartheta}) = \sum_{i=1}^K w_i \frac{\mathrm{KL}(P_{\boldsymbol{\theta},i}\|P_{\boldsymbol{\vartheta},i})}{C_i(\boldsymbol{\theta})}, \qquad D_x^w(\boldsymbol{w};\boldsymbol{\theta}) = \inf_{\boldsymbol{\vartheta}\in\mathrm{Alt}_x(\boldsymbol{\theta})} \tilde{\Gamma}^c(\boldsymbol{w},\boldsymbol{\vartheta}).$$

Since Θ is compact (Assumption 1), the feasible set $\mathrm{cl}(\mathrm{Alt}_x(\boldsymbol{\theta})) \subset \Theta$ is compact. For each i, the map $\boldsymbol{\vartheta} \mapsto \mathrm{KL}(P_{\boldsymbol{\theta},i}\|P_{\boldsymbol{\vartheta},i})$ is continuous on Θ (one-dimensional canonical exponential family with continuous η_i and $A \in C^2$ on a neighborhood of $\eta_i(\Theta)$). Hence $\boldsymbol{\vartheta} \mapsto \tilde{\Gamma}^c(\boldsymbol{w},\boldsymbol{\vartheta})$ is continuous on $\mathrm{cl}(\mathrm{Alt}_x(\boldsymbol{\theta}))$ for each fixed $\boldsymbol{w}$, and therefore

$$D_x^w(\boldsymbol{w};\boldsymbol{\theta}) = \min_{\boldsymbol{\vartheta}\in\mathrm{cl}(\mathrm{Alt}_x(\boldsymbol{\theta}))} \tilde{\Gamma}^c(\boldsymbol{w},\boldsymbol{\vartheta}).$$

Shape on $\mathbb{R}^K_{\ge 0}$. Fix any $\boldsymbol{\vartheta}$. The map $\boldsymbol{w} \mapsto \tilde{\Gamma}^c(\boldsymbol{w},\boldsymbol{\vartheta})$ is linear with nonnegative coefficients (since $\mathrm{KL} \ge 0$ and $C_i(\boldsymbol{\theta}) \ge c_{\min} > 0$ by Assumption 2). Thus it is concave, continuous, coordinatewise nondecreasing, and degree-one homogeneous on $\mathbb{R}^K_{\ge 0}$. Taking the minimum over $\boldsymbol{\vartheta} \in \mathrm{cl}(\mathrm{Alt}_x(\boldsymbol{\theta}))$ preserves concavity, monotonicity, homogeneity, and nonnegativity, so $D_x^w(\cdot;\boldsymbol{\theta})$ has the claimed properties.

Unique minimizer and C^1 smoothness on $\mathbb{R}^K_{>0}$. Fix $\boldsymbol{w} \in \mathbb{R}^K_{>0}$ so that $w_i > 0$ for all i. Existence of a minimizer follows from compactness of $\mathrm{cl}(\mathrm{Alt}_x(\boldsymbol{\theta}))$.

For uniqueness, note that for a one-dimensional canonical exponential family the KL divergence is a Bregman divergence of A and is strictly convex in the natural parameter; under our structured models (e.g., linear bandits where η_i is affine in $\boldsymbol{\vartheta}$), each map $\boldsymbol{\vartheta} \mapsto \mathrm{KL}(P_{\boldsymbol{\theta},i} \| P_{\boldsymbol{\vartheta},i})$ is convex and not affine on any nontrivial segment. Since $w_i/C_i(\boldsymbol{\theta}) > 0$ for all i and $\mathrm{cl}(\mathrm{Alt}_x(\boldsymbol{\theta}))$ is convex (Assumption 5), the weighted sum $\boldsymbol{\vartheta} \mapsto \tilde{\Gamma}^c(\boldsymbol{w}, \boldsymbol{\vartheta})$ is strictly convex on $\mathrm{cl}(\mathrm{Alt}_x(\boldsymbol{\theta}))$, hence the minimizer is unique; denote it by $\boldsymbol{\vartheta}_x(\boldsymbol{w}; \boldsymbol{\theta})$.

By Danskin's theorem, $D^w_x(\cdot; \boldsymbol{\theta})$ is differentiable at $\boldsymbol{w}$ and

$$\big[\nabla_{\boldsymbol{w}} D^w_x(\boldsymbol{w}; \boldsymbol{\theta})\big]_i = \frac{\partial}{\partial w_i} \tilde{\Gamma}^c(\boldsymbol{w}, \boldsymbol{\vartheta}_x(\boldsymbol{w}; \boldsymbol{\theta})) = \frac{\mathrm{KL}\big(P_{\boldsymbol{\theta},i} \| P_{\boldsymbol{\vartheta}_x(\boldsymbol{w};\boldsymbol{\theta}),i}\big)}{C_i(\boldsymbol{\theta})}, \qquad i \in [K].$$

Moreover, the argmin map $\boldsymbol{w} \mapsto \boldsymbol{\vartheta}_x(\boldsymbol{w}; \boldsymbol{\theta})$ is continuous on $\mathbb{R}^K_{>0}$ by Berge's maximum theorem (constant compact feasible set and unique minimizer), hence the displayed gradient is continuous in $\boldsymbol{w}$. Therefore $D^w_x(\cdot; \boldsymbol{\theta}) \in C^1(\mathbb{R}^K_{>0})$.

Gradient lower bound and strict positivity on $\mathrm{int}(\Delta_K)$. Assume $D^w_x(\cdot; \boldsymbol{\theta}) \not\equiv 0$ on Δ_K. Consider the continuous function on the compact set $\mathrm{cl}(\mathrm{Alt}_x(\boldsymbol{\theta}))$,

$$\boldsymbol{\vartheta} \mapsto \max_{i \in [K]} \frac{\mathrm{KL}\big(P_{\boldsymbol{\theta},i} \| P_{\boldsymbol{\vartheta},i}\big)}{c_{\max}}.$$

If this maximum equals 0 at some $\boldsymbol{\vartheta}$, then $\mathrm{KL}(P_{\boldsymbol{\theta},i} \| P_{\boldsymbol{\vartheta},i}) = 0$ for all i, hence $P_{\boldsymbol{\theta},i} = P_{\boldsymbol{\vartheta},i}$ for all i. By the identifiability condition in Assumption 3, this implies $\mathcal{I}(\boldsymbol{\vartheta}) = \mathcal{I}(\boldsymbol{\theta})$, contradicting $\boldsymbol{\vartheta} \in \mathrm{cl}(\mathrm{Alt}_x(\boldsymbol{\theta})) \subset \mathrm{Alt}(\boldsymbol{\theta})$. Therefore the maximum is strictly positive on $\mathrm{cl}(\mathrm{Alt}_x(\boldsymbol{\theta}))$, and by compactness,

$$d_x(\boldsymbol{\theta}) \triangleq \min_{\boldsymbol{\vartheta} \in \mathrm{cl}(\mathrm{Alt}_x(\boldsymbol{\theta}))} \max_{i \in [K]} \frac{\mathrm{KL}\big(P_{\boldsymbol{\theta},i} \| P_{\boldsymbol{\vartheta},i}\big)}{c_{\max}} > 0.$$

Now fix any $\boldsymbol{w} \in \mathrm{int}(\Delta_K)$. Using the gradient formula and $C_i(\boldsymbol{\theta}) \le c_{\max}$,

$$\sum_{i=1}^K \big[\nabla_{\boldsymbol{w}} D^w_x(\boldsymbol{w}; \boldsymbol{\theta})\big]_i = \sum_{i=1}^K \frac{\mathrm{KL}\big(P_{\boldsymbol{\theta},i} \| P_{\boldsymbol{\vartheta}_x(\boldsymbol{w};\boldsymbol{\theta}),i}\big)}{C_i(\boldsymbol{\theta})} \ge \frac{1}{c_{\max}} \max_{i \in [K]} \mathrm{KL}\big(P_{\boldsymbol{\theta},i} \| P_{\boldsymbol{\vartheta}_x(\boldsymbol{w};\boldsymbol{\theta}),i}\big) \ge d_x(\boldsymbol{\theta}).$$

Finally, since $D^w_x(\cdot; \boldsymbol{\theta})$ is degree-one homogeneous and differentiable at $\boldsymbol{w} \in \mathrm{int}(\Delta_K)$, Euler's identity yields

$$D^w_x(\boldsymbol{w}; \boldsymbol{\theta}) = \langle \boldsymbol{w}, \nabla_{\boldsymbol{w}} D^w_x(\boldsymbol{w}; \boldsymbol{\theta}) \rangle \ge \Big(\min_{i \in [K]} w_i\Big) \sum_{i=1}^K \big[\nabla_{\boldsymbol{w}} D^w_x(\boldsymbol{w}; \boldsymbol{\theta})\big]_i \ge \Big(\min_{i \in [K]} w_i\Big) d_x(\boldsymbol{\theta}) > 0,$$

so $D^w_x(\cdot; \boldsymbol{\theta})$ is strictly positive on $\mathrm{int}(\Delta_K)$.

C Proof of Lemma 2

Detection consistency follows directly from (12), the definition (10) of D_x^w and the continuity of $C_i(\cdot)$ (assumption 2). For selection consistency on non-forced rounds, note that $q_{n,i} = H_i^{x_n}(\boldsymbol{p}_n; \boldsymbol{\theta}_n)$ and use the relation (14),

$$\widetilde{h}_{n,i} = \frac{q_{n,i}\frac{C_i(\boldsymbol{\theta})}{B_n+C_i(\boldsymbol{\theta})}}{\sum_{j=1}^K q_{n,j}\frac{C_j(\boldsymbol{\theta})}{B_n+C_j(\boldsymbol{\theta})}} = \frac{H_i^{x_n}(\boldsymbol{p}_n;\boldsymbol{\theta}_n)\cdot\frac{C_i(\boldsymbol{\theta})}{B_n+C_j(\boldsymbol{\theta})}}{\sum_{j=1}^K H_j^{x_n}(\boldsymbol{p}_n;\boldsymbol{\theta}_n)\cdot\frac{C_j(\boldsymbol{\theta})}{B_n+C_j(\boldsymbol{\theta})}} = \frac{h_i^{x_n}(\boldsymbol{w}_n;\boldsymbol{\theta}_n)\cdot\frac{1}{B_n+C_j(\boldsymbol{\theta})}}{\sum_{j=1}^K h_j^{x_n}(\boldsymbol{w}_n;\boldsymbol{\theta}_n)\cdot\frac{1}{B_n+C_j(\boldsymbol{\theta})}}$$
$$= \frac{h_i^{x_n}(\boldsymbol{w}_n;\boldsymbol{\theta}_n)\cdot\frac{B_n}{B_n+C_i(\boldsymbol{\theta})}}{\sum_{j=1}^K h_j^{x_n}(\boldsymbol{w}_n;\boldsymbol{\theta}_n)\cdot\frac{B_n}{B_n+C_j(\boldsymbol{\theta})}} = \frac{h_i^{x_n}(\boldsymbol{w}_n;\boldsymbol{\theta}_n)\cdot(1+x_i)^{-1}}{\sum_{j=1}^K h_j^{x_n}(\boldsymbol{w}_n;\boldsymbol{\theta}_n)\cdot(1+x_j)^{-1}}$$

where $x_i \triangleq C_i(\boldsymbol{\theta})/B_n \in (\frac{c_{\min}}{nc_{\max}}, \frac{c_{\max}}{nc_{\min}})$ then by $1-x \le (1+x)^{-1} \le 1$,

$$\widetilde{h}_{n,i} \ge \frac{h_i^{x_n}(\boldsymbol{w}_n;\boldsymbol{\theta}_n)\cdot(1-x_i)}{\sum h_j^{x_n}(\boldsymbol{w}_n;\boldsymbol{\theta}_n)} \ge (1-x_i)h_i^{x_n}(\boldsymbol{w}_n;\boldsymbol{\theta}_n) \ge \left(1-\frac{c_{\max}}{nc_{\min}}\right)h_i^{x_n}(\boldsymbol{w}_n;\boldsymbol{\theta}_n),$$
$$\widetilde{h}_{n,i} \le \frac{h_i^{x_n}(\boldsymbol{w}_n;\boldsymbol{\theta}_n)}{\sum h_j^{x_n}(\boldsymbol{w}_n;\boldsymbol{\theta}_n)\cdot(1+x_j)^{-1}} \le \left(1+\frac{c_{\max}}{nc_{\min}}\right)h_i^{x_n}(\boldsymbol{w}_n;\boldsymbol{\theta}_n).$$

Thus

$$\left|\widetilde{h}_{n,i} - h_i^{x_n}(\boldsymbol{w}_n,\boldsymbol{\theta}_n)\right| \le \frac{c_{\max}}{nc_{\min}}h_i^{x_n}(\boldsymbol{w}_n,\boldsymbol{\theta}_n) \le \frac{c_{\max}}{nc_{\min}} \to 0.$$

D Proof of Proposition 1

Properties of `det`. Fix $\boldsymbol{w} \in \Delta_K$. Since $F(\boldsymbol{w},\cdot)$ is linear on the compact convex set $\Delta_{|\mathcal{X}|}$, the minimizer set $\texttt{det}(\boldsymbol{w}) = \arg\min_{\boldsymbol{\nu}\in\Delta_{|\mathcal{X}|}} F(\boldsymbol{w},\boldsymbol{\nu})$ is nonempty and is a face of $\Delta_{|\mathcal{X}|}$, hence convex and compact. Because $\mathcal{X}$ is finite and each $D_x^w(\boldsymbol{w})$ is continuous in $\boldsymbol{w}$, the function $F(\boldsymbol{w},\boldsymbol{\nu}) = \sum_{x\in\mathcal{X}}\nu_x D_x^w(\boldsymbol{w})$ is jointly continuous on $\Delta_K \times \Delta_{|\mathcal{X}|}$. By Berge's maximum theorem, `det` is upper hemicontinuous and compact-valued; in particular, its graph is closed.

The limit-set $S(\boldsymbol{w})$. For each $\boldsymbol{w} \in \Delta_K$, define

$$S(\boldsymbol{w}) \triangleq \left\{\lim_{n\to\infty}\boldsymbol{h}(\boldsymbol{w}^n,\boldsymbol{\nu}^n) :\ \boldsymbol{w}^n \to \boldsymbol{w},\ \boldsymbol{w}^n \in \mathcal{R},\ \boldsymbol{\nu}^n \in \texttt{det}(\boldsymbol{w}^n)\right\} \subseteq \Delta_K.$$

Nonemptiness. Pick any sequence $\boldsymbol{w}^n \in \mathcal{R}$ with $\boldsymbol{w}^n \to \boldsymbol{w}$. For each n, choose $\boldsymbol{\nu}^n \in \texttt{det}(\boldsymbol{w}^n)$. Then $\boldsymbol{h}(\boldsymbol{w}^n,\boldsymbol{\nu}^n) \in \Delta_K$, and by compactness of Δ_K the sequence admits a convergent subsequence. Hence $S(\boldsymbol{w}) \neq \emptyset$.

Compactness. Since $S(\boldsymbol{w}) \subseteq \Delta_K$ and Δ_K is compact, it suffices to show that $S(\boldsymbol{w})$ is closed. Take any sequence $\boldsymbol{s}^m \in S(\boldsymbol{w})$ with $\boldsymbol{s}^m \to \boldsymbol{s}$. For each m, by definition of $S(\boldsymbol{w})$ there exist sequences $\{\boldsymbol{w}^{m,n}\}_{n\ge1} \subset \mathcal{R}$ and $\{\boldsymbol{\nu}^{m,n}\}_{n\ge1}$ such that

$$\boldsymbol{w}^{m,n} \to \boldsymbol{w}, \qquad \boldsymbol{\nu}^{m,n} \in \texttt{det}(\boldsymbol{w}^{m,n}), \qquad \boldsymbol{h}(\boldsymbol{w}^{m,n},\boldsymbol{\nu}^{m,n}) \to \boldsymbol{s}^m \quad \text{as } n\to\infty.$$

Choose an ε-selection: pick $n(m)$ large enough so that

$$\|\boldsymbol{w}^{m,n(m)} - \boldsymbol{w}\| \le \frac{1}{m}, \qquad \|\boldsymbol{h}(\boldsymbol{w}^{m,n(m)},\boldsymbol{\nu}^{m,n(m)}) - \boldsymbol{s}^m\| \le \frac{1}{m}.$$

Set $\widetilde{\boldsymbol{w}}^m := \boldsymbol{w}^{m,n(m)} \in \mathcal{R}$, $\widetilde{\boldsymbol{\nu}}^m := \boldsymbol{\nu}^{m,n(m)} \in \texttt{det}(\widetilde{\boldsymbol{w}}^m)$, and $\widetilde{\boldsymbol{s}}^m := \boldsymbol{h}(\widetilde{\boldsymbol{w}}^m, \widetilde{\boldsymbol{\nu}}^m) \in \Delta_K$. Then $\widetilde{\boldsymbol{w}}^m \to \boldsymbol{w}$ and $\|\widetilde{\boldsymbol{s}}^m - \boldsymbol{s}^m\| \le 1/m$, hence $\widetilde{\boldsymbol{s}}^m \to \boldsymbol{s}$.

By compactness of $\Delta_{|\mathcal{X}|}$, along a subsequence (not relabeled) $\widetilde{\boldsymbol{\nu}}^m \to \boldsymbol{\nu}$. Since $\widetilde{\boldsymbol{w}}^m \to \boldsymbol{w}$ and $\widetilde{\boldsymbol{\nu}}^m \in \texttt{det}(\widetilde{\boldsymbol{w}}^m)$ with $\widetilde{\boldsymbol{\nu}}^m \to \boldsymbol{\nu}$, the closed-graph property of $\texttt{det}$ implies $\boldsymbol{\nu} \in \texttt{det}(\boldsymbol{w})$. Moreover, by construction,

$$\widetilde{\boldsymbol{s}}^m = \boldsymbol{h}(\widetilde{\boldsymbol{w}}^m, \widetilde{\boldsymbol{\nu}}^m) \to \boldsymbol{s} \quad \text{with} \quad \widetilde{\boldsymbol{w}}^m \in \mathcal{R},\ \widetilde{\boldsymbol{w}}^m \to \boldsymbol{w},\ \widetilde{\boldsymbol{\nu}}^m \in \texttt{det}(\widetilde{\boldsymbol{w}}^m),$$

so $\boldsymbol{s} \in S(\boldsymbol{w})$. Thus $S(\boldsymbol{w})$ is closed, hence compact.

Upper hemicontinuity of $S(\cdot)$. We show graph(S) is closed. Let $\boldsymbol{w}^m \to \boldsymbol{w}$ and take $\boldsymbol{s}^m \in S(\boldsymbol{w}^m)$ with $\boldsymbol{s}^m \to \boldsymbol{s}$. For each m, pick sequences $\{\boldsymbol{w}^{m,n}\}_{n\ge1} \subset \mathcal{R}$ and $\{\boldsymbol{\nu}^{m,n}\}_{n\ge1}$ such that

$$\boldsymbol{w}^{m,n} \to \boldsymbol{w}^m, \qquad \boldsymbol{\nu}^{m,n} \in \texttt{det}(\boldsymbol{w}^{m,n}), \qquad \boldsymbol{h}(\boldsymbol{w}^{m,n}, \boldsymbol{\nu}^{m,n}) \to \boldsymbol{s}^m \quad (n \to \infty).$$

As above, choose $n(m)$ so that

$$\|\boldsymbol{w}^{m,n(m)} - \boldsymbol{w}^m\| \le \frac{1}{m}, \qquad \|\boldsymbol{h}(\boldsymbol{w}^{m,n(m)}, \boldsymbol{\nu}^{m,n(m)}) - \boldsymbol{s}^m\| \le \frac{1}{m}.$$

Define $\widetilde{\boldsymbol{w}}^m := \boldsymbol{w}^{m,n(m)} \in \mathcal{R}$, $\widetilde{\boldsymbol{\nu}}^m := \boldsymbol{\nu}^{m,n(m)} \in \texttt{det}(\widetilde{\boldsymbol{w}}^m)$, and $\widetilde{\boldsymbol{s}}^m := \boldsymbol{h}(\widetilde{\boldsymbol{w}}^m, \widetilde{\boldsymbol{\nu}}^m) \in \Delta_K$. Then

$$\|\widetilde{\boldsymbol{w}}^m - \boldsymbol{w}\| \le \|\widetilde{\boldsymbol{w}}^m - \boldsymbol{w}^m\| + \|\boldsymbol{w}^m - \boldsymbol{w}\| \to 0, \qquad \|\widetilde{\boldsymbol{s}}^m - \boldsymbol{s}\| \le \|\widetilde{\boldsymbol{s}}^m - \boldsymbol{s}^m\| + \|\boldsymbol{s}^m - \boldsymbol{s}\| \to 0,$$

so $\widetilde{\boldsymbol{w}}^m \to \boldsymbol{w}$ and $\widetilde{\boldsymbol{s}}^m \to \boldsymbol{s}$. By compactness of $\Delta_{|\mathcal{X}|}$, along a subsequence $\widetilde{\boldsymbol{\nu}}^m \to \boldsymbol{\nu}$. Using again the closed graph of $\texttt{det}$, we get $\boldsymbol{\nu} \in \texttt{det}(\boldsymbol{w})$. Therefore $\boldsymbol{s} = \lim_m \boldsymbol{h}(\widetilde{\boldsymbol{w}}^m, \widetilde{\boldsymbol{\nu}}^m) \in S(\boldsymbol{w})$, proving the closed-graph property. Since S is compact-valued (Step 2), it follows that S is upper hemicontinuous.

The correspondence $\texttt{sel}$ inherits regularity. By definition, $\texttt{sel}(\boldsymbol{w}) = \text{conv}(S(\boldsymbol{w}))$. Since $S(\boldsymbol{w})$ is nonempty and compact, $\texttt{sel}(\boldsymbol{w})$ is nonempty, convex, and compact.

It remains to show u.h.c. of $\texttt{sel}$. We use a closed-graph argument and Carathéodory. Let $\boldsymbol{w}^m \to \boldsymbol{w}$ and $\boldsymbol{s}^m \in \texttt{sel}(\boldsymbol{w}^m)$ with $\boldsymbol{s}^m \to \boldsymbol{s}$. By Carathéodory in $\mathbb{R}^K$, for each m we may write

$$\boldsymbol{s}^m = \sum_{j=1}^{K} \lambda^{m,j} \boldsymbol{x}^{m,j}, \qquad \lambda^{m,j} \ge 0, \quad \sum_{j=1}^{K} \lambda^{m,j} = 1, \quad \boldsymbol{x}^{m,j} \in S(\boldsymbol{w}^m).$$

By compactness of Δ_K and Δ_K (for the coefficients), after passing to a subsequence we may assume

$$\lambda^{m,j} \to \lambda^j, \qquad \boldsymbol{x}^{m,j} \to \boldsymbol{x}^j \quad \text{for each } j = 1, \dots, K.$$

Since S is u.h.c. with compact values, its graph is closed; thus $\boldsymbol{x}^{m,j} \in S(\boldsymbol{w}^m)$ and $(\boldsymbol{w}^m, \boldsymbol{x}^{m,j}) \to (\boldsymbol{w}, \boldsymbol{x}^j)$ imply $\boldsymbol{x}^j \in S(\boldsymbol{w})$. Taking limits in the convex combination yields

$$\boldsymbol{s} = \lim_{m\to\infty} \boldsymbol{s}^m = \sum_{j=1}^{K} \lambda^j \boldsymbol{x}^j \in \text{conv}(S(\boldsymbol{w})) = \texttt{sel}(\boldsymbol{w}).$$

Hence graph$(\texttt{sel})$ is closed. Because $\texttt{sel}$ is compact-valued, it is upper hemicontinuous.

Existence of DI solutions and forward invariance. Because $\texttt{sel}$ is nonempty, convex, compact-valued, and u.h.c., the same holds for $G(\boldsymbol{w}) = \texttt{sel}(\boldsymbol{w}) - \boldsymbol{w}$. Also, for any $\boldsymbol{w} \in \Delta_K$ and $\boldsymbol{g} \in G(\boldsymbol{w})$, we can write $\boldsymbol{g} = \boldsymbol{s} - \boldsymbol{w}$ with $\boldsymbol{s} \in \texttt{sel}(\boldsymbol{w}) \subseteq \Delta_K$, hence $\|\boldsymbol{g}\|_1 \leq 2$; thus G is bounded. Standard existence theorems for differential inclusions with u.h.c. nonempty convex compact right-hand sides (e.g. Aubin and Cellina 1984) yield an absolutely continuous solution to (18) for any initial condition $\boldsymbol{w}_0 \in \Delta_K$.

Moreover, Δ_K is forward invariant: if $\dot{\boldsymbol{w}}(t) = \boldsymbol{s}(t) - \boldsymbol{w}(t)$ with $\boldsymbol{s}(t) \in \texttt{sel}(\boldsymbol{w}(t)) \subseteq \Delta_K$ a.e., then $\sum_{i=1}^K \dot{w}_i(t) = 1 - 1 = 0$ a.e., and if $p_i(t) = 0$ then $\dot{w}_i(t) = s_i(t) \geq 0$ a.e. Hence $\boldsymbol{w}(t) \in \Delta_K$ for all $t \geq 0$.

Finally, for any absolutely continuous solution $\boldsymbol{w}(\cdot)$, define $\boldsymbol{s}(t) := \dot{\boldsymbol{w}}(t) + \boldsymbol{w}(t)$ for a.e. t. Then $\boldsymbol{s}(\cdot)$ is measurable and satisfies $\boldsymbol{s}(t) \in \texttt{sel}(\boldsymbol{w}(t))$ for a.e. t.

E Proof of Theorem 3

Our analysis proceeds in four steps. First, in Appendix E.1, we construct a globally well-defined information value that serves as a Lyapunov function for the limiting differential inclusion and show that it is nondecreasing along every solution trajectory. By LaSalle's invariance principle, the associated ω-limit set has empty interior. Second, in Appendix E.2, we transfer this continuous-time structure to the stochastic iterates via stochastic-approximation tools—the asymptotic pseudo-trajectory (APT) framework and the characterization of internally chain transitive (ICT) sets—which yields convergence of the information value along the discrete-time trajectory. Third, in Appendix E.3, we rule out convergence to the degenerate value 0 under interior initialization, ensuring the dynamics remains in a positive-information regime. Finally, in Appendix E.4, we use the Kullback–Leibler divergence as an energy function on the active set and establish a uniform negative-drift bound. This forces the limiting information value to coincide with the optimal value F^*, completing the proof.

E.1 Continuous-Time Dynamics

This subsection analyzes the limiting differential inclusion (DI) by constructing a global (weak) Lyapunov function and characterizing its limit behavior. Let

$$V(t) = \min_{x \in \mathcal{X}} D_x^w(\boldsymbol{w}(t)) \tag{34}$$

along any absolutely continuous DI solution $\boldsymbol{w}(\cdot)$. We first show that $\boldsymbol{w} \mapsto \min_{x \in \mathcal{X}} D_x^w(\boldsymbol{w})$ is Lipschitz on Δ_K (Lemma 7), which implies $V(\cdot)$ is absolutely continuous along DI trajectories. Because V is generally nonsmooth, we then establish a Clarke-type chain rule along absolutely continuous curves (Lemma 8), enabling differentiation of $V(t)$ almost everywhere.

Using these tools, we prove that $V(t)$ is nondecreasing along every DI solution (Theorem 4), so V serves as a global Lyapunov function. Since this Lyapunov function is not strict and the DI may admit multiple equilibria, we invoke LaSalle's invariance principle to characterize ω–limit

points. In particular, any ω–limit point $\boldsymbol{w}^*$ is facewise optimal: writing $J = \mathrm{Supp}(\boldsymbol{w}^*)$, it maximizes $\min_{x\in\mathcal{X}} D_x^w(\boldsymbol{w})$ over the face Δ_J (Corollary 1). As a consequence, the set of stationary objective values is

$$\left\{ \max_{\boldsymbol{w}\in\Delta_J} \min_{x\in\mathcal{X}} D_x^w(\boldsymbol{w}) \ : \ \emptyset \neq J \subseteq [K] \right\},$$

which is finite (hence discrete and with empty interior). This finiteness yields the *weak Sard property* used later in the stochastic-approximation analysis of the discrete-time iterates.

Lemma 7. *Under Assumption 7, the function $\boldsymbol{w} \mapsto \min_{x\in\mathcal{X}} D_x^w(\boldsymbol{w})$ is M-Lipschitz on Δ_K with respect to $\|\cdot\|_1$, i.e.,*

$$\left|\min_{x\in\mathcal{X}} D_x^w(\boldsymbol{w}_1) - \min_{x\in\mathcal{X}} D_x^w(\boldsymbol{w}_2)\right| \leq M\|\boldsymbol{w}_1 - \boldsymbol{w}_2\|_1, \qquad \forall \boldsymbol{w}_1, \boldsymbol{w}_2 \in \Delta_K.$$

Proof. By Assumption 7, each D_x is M-Lipschitz on Δ_K with respect to $\|\cdot\|_1$. Fix $\boldsymbol{w}_1, \boldsymbol{w}_2 \in \Delta_K$ and choose $x_2 \in \arg\min_{x\in\mathcal{X}} D_x^w(\boldsymbol{w}_2)$. Then

$$\min_{x\in\mathcal{X}} D_x^w(\boldsymbol{w}_1) - \min_{x\in\mathcal{X}} D_x^w(\boldsymbol{w}_2) \leq D_{x_2}^w(\boldsymbol{w}_1) - D_{x_2}^w(\boldsymbol{w}_2) \leq M\|\boldsymbol{w}_1 - \boldsymbol{w}_2\|_1.$$

Swapping $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ yields the reverse inequality, proving the claim. □

Lemma 8 (Chain rule for $\min_{x\in\mathcal{X}} D_x^w(\boldsymbol{w})$)**.** *The information function $\min_{x\in\mathcal{X}} D_x^w(\boldsymbol{w})$ admits the Clarke chain rule: for any absolutely continuous curve $\boldsymbol{p}(t) : [0,\infty) \to \Delta_K$, $\min_{x\in\mathcal{X}} D_x^w(\boldsymbol{w}(t))$ is differentiable for almost every $t \in [0,\infty)$ and*

$$\frac{\mathrm{d}}{\mathrm{d}t} \min_{x\in\mathcal{X}} D_x^w(\boldsymbol{w}(t)) \ = \ \boldsymbol{f}(\boldsymbol{w}(t))^\top \frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{w}(t) \quad \text{for every } \boldsymbol{f}(\boldsymbol{w}(t)) \in \partial^\circ \min_{x\in\mathcal{X}} D_x^w(\boldsymbol{w}(t)), \tag{35}$$

for almost every t.

Proof. Since $\min_{x\in\mathcal{X}} D_x^w(\boldsymbol{w})$ is concave (Lemma 1) and locally Lipschitz on Δ_K (Lemma 7), the chain rule for concave function along absolutely continuous curves (see, e.g., Brezis (1973)) yields the claim. □

Remark 1 (More general chain-rule conditions)**.** *Lemma 8 is a special case of general chain rules for subdifferentially regular functions, Davis et al. (2020, Lemma 5.2) and Bolte and Pauwels (2021, Section 3.2) gives a broader characterization, known as path differentiable functions.*

The main result for this section is the convergence of the continuous-time dynamics to stationary points via Lyapunov analysis. Recall the information value along the DI solution $\boldsymbol{w}(\cdot)$ defined in (34). We show that such a information value *does not decrease*, and serve as a (weak) Lyapunov function.

Theorem 4 (Global Lyapunov function)**.** *Let $\boldsymbol{w}(\cdot)$ be any absolutely continuous solution of* (18)*, then $V(\cdot)$ is absolutely continuous and*

$$\frac{\mathrm{d}}{\mathrm{d}t} V(t) \ \geq \ 0 \quad \text{for a.e. } t \geq 0.$$

Proof. By the definition of the selection correspondence and Carathéodory's theorem, there exist sequences

$$(\boldsymbol{w}^{n,j}, \boldsymbol{\nu}^{n,j})_{n\to\infty} \subseteq \mathcal{R} \times \Delta_{|\mathcal{X}|}, \qquad j \in [K],$$

and weights $(\lambda^j)_{j\in[K]} \in \Delta_K$ such that $\boldsymbol{w}^{n,j} \xrightarrow{n\to\infty} \boldsymbol{w}(t)$ for each j and

$$\sum_{j\in[K]} \lambda^j \boldsymbol{h}(\boldsymbol{w}^{n,j}, \boldsymbol{\nu}^{n,j}) - \boldsymbol{w}(t) \xrightarrow{n\to\infty} \frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{w}(t). \tag{36}$$

The limit (upon possibly passing to a subsequence)

$$\boldsymbol{f}^*(\boldsymbol{w}(t)) \triangleq \lim_{n\to\infty} \sum_{j\in[K]} \lambda^j \nabla_{\boldsymbol{w}} F(\boldsymbol{w}^{n,j}, \boldsymbol{\nu}^{n,j}) \in \partial^\circ \min_{x\in\mathcal{X}} D_x^w(\boldsymbol{w}(t)) \tag{37}$$

since the Clarke subdifferential is closed under such limits. Since $\boldsymbol{w}(\cdot)$ is absolutely continuous and $\boldsymbol{w} \mapsto \min_{x\in\mathcal{X}} D_x^w(\boldsymbol{w})$ is Lipschitz on Δ_K (Lemma 7), the map $V(t) = \min_{x\in\mathcal{X}} D_x^w(\boldsymbol{w}(t))$ is absolutely continuous and hence differentiable for a.e. $t \geq 0$ and lemma 8 gives

$$\frac{\mathrm{d}}{\mathrm{d}t}V(t) = \boldsymbol{f}^*(\boldsymbol{w}(t))^\top \frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{w}(t), \quad \text{for a.e. } t.$$

We claim that this time derivative is nonnegative, which by (36) and (37) is sufficient to show

$$\lim_{n\to\infty} \left(\sum_{j\in[K]} \lambda^j \nabla_{\boldsymbol{w}} F(\boldsymbol{w}^{n,j}, \boldsymbol{\nu}^{n,j})\right)^\top \left(\sum_{j\in[K]} \lambda^j \boldsymbol{h}(\boldsymbol{w}^{n,j}, \boldsymbol{\nu}^{n,j})\right) \;\geq\; \boldsymbol{f}^*(\boldsymbol{w}(t)) \circ \boldsymbol{w}(t).$$

Indeed, for any fixed n, introduce the shorthand

$$w_i^j \triangleq w_i^{n,j}, \qquad f_i^j \triangleq \lambda^j \big[\nabla_{\boldsymbol{w}} F(\boldsymbol{w}^{n,j}, \boldsymbol{\nu}^{n,j})\big]_i, \qquad F^j \triangleq \sum_{i\in[K]} w_i^j f_i^j = \lambda^j F(\boldsymbol{w}^{n,j}, \boldsymbol{\nu}^{n,j}).$$

Then the weighted IDS probability satisfies $\lambda^j h_i(\boldsymbol{w}^{n,j}, \boldsymbol{\nu}^{n,j}) = \frac{w_i^j f_i^j}{F^j}$, and we can compute

$$\begin{aligned}
&\left(\sum_{j\in[K]} \lambda^j \nabla_{\boldsymbol{w}} F(\boldsymbol{w}^{n,j}, \boldsymbol{\nu}^{n,j})\right)^\top \left(\sum_{j\in[K]} \lambda^j \boldsymbol{h}(\boldsymbol{w}^{n,j}, \boldsymbol{\nu}^{n,j})\right) \\
&= \sum_{i\in[K]} \left(\sum_{j\in[K]} f_i^j\right) \left(\sum_{j\in[K]} \frac{w_i^j f_i^j}{F^j}\right) \\
&= \sum_{i\in[K]} \sum_{j\in[K]} \frac{w_i^j f_i^j}{F^j} \Big(\sum_{k\in[K]} f_i^k\Big) \\
&= \sum_{j\in[K]} \frac{1}{F^j} \sum_{i\in[K]} w_i^j f_i^j \Big(\sum_{k\in[K]} f_i^k\Big),
\end{aligned} \tag{38}$$

For each fixed $j \in [K]$, using $\sum_{k\in[K]} f_i^k \geq f_i^j$ and the Cauchy–Schwarz inequality, we have

$$\begin{aligned}
\sum_{i\in[K]} w_i^j f_i^j \Big(\sum_{k\in[K]} f_i^k\Big) &\geq \sum_{i\in[K]} w_i^j (f_i^j)^2 = \Big(\sum_{i\in[K]} w_i^j (f_i^j)^2\Big)\Big(\sum_{i\in[K]} w_i^j\Big) \\
&\geq \Big(\sum_{i\in[K]} w_i^j f_i^j\Big)^2 = (F^j)^2.
\end{aligned} \tag{39}$$

Summing over j and using (38), we obtain

$$(38) \geq \sum_{j\in[K]} \frac{1}{F^j}(F^j)^2 = \sum_{j\in[K]} F^j = \sum_{j\in[K]} \lambda^j F(\boldsymbol{w}^{n,j}, \boldsymbol{\nu}^{n,j}) \xrightarrow{n\to\infty} \min_{x\in\mathcal{X}} D_x^w(\boldsymbol{w}(t)).$$

An application of the Euler identity for the Clarke gradient gives

$$\boldsymbol{f}^*(\boldsymbol{w}(t))^\top \boldsymbol{w}(t) = \min_{x\in\mathcal{X}} D_x^w(\boldsymbol{w}(t)),$$

which establishes the desired claim and

$$\frac{\mathrm{d}}{\mathrm{d}t}V(t) = \boldsymbol{f}^*(\boldsymbol{w}(t))^\top \frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{w}(t) \geq 0 \quad \text{for a.e. } t.$$

□

Corollary 1 (LaSalle limit points and stationary values)**.** *Let $\boldsymbol{w}^* \in \Delta_K$ be any ω–limit point of the DI* (18)*. Then there exists $\boldsymbol{\nu}^* \in \mathtt{det}(\boldsymbol{w}^*)$ and $\boldsymbol{f}(\boldsymbol{w}^*, \boldsymbol{\nu}^*) \in \partial^\circ \min_{x\in\mathcal{X}} D_x^w(\boldsymbol{w}(t))$ such that*

$$\big[\boldsymbol{f}(\boldsymbol{w}^*, \boldsymbol{\nu}^*)\big]_i = F(\boldsymbol{w}^*, \boldsymbol{\nu}^*) \quad \textit{for all } i \in \mathrm{Supp}(\boldsymbol{w}^*). \tag{40}$$

Writing $J \triangleq \mathrm{Supp}(\boldsymbol{w}^*)$ and the face $\Delta_J \triangleq \{\boldsymbol{w} \in \Delta_K : p_i = 0 \ (i \notin J)\}$, we have the facewise maximality

$$\min_{x\in\mathcal{X}} D_x^w(\boldsymbol{w}^*) = \max_{\boldsymbol{w}\in\Delta_J} \min_{x\in\mathcal{X}} D_x^w(\boldsymbol{w}).$$

Consequently, the set of all stationary values is

$$\Big\{ \max_{\boldsymbol{w}\in\Delta_J} \min_{x\in\mathcal{X}} D_x^w(\boldsymbol{w}) \ : \emptyset \neq J \subseteq [K] \Big\},$$

which is finite, hence has empty interior.

Proof. By Theorem 4, $t \mapsto \min_x D_x^w(\boldsymbol{w}(t))$ is nondecreasing. LaSalle's invariance principle for DIs implies that every ω–limit point lies in the largest weakly invariant subset of the set where the time derivative of $t \mapsto \min_x D_x^w(\boldsymbol{w}(t))$ vanishes.

Take limit to the derivative identity (39) where the Cauchy–Schwarz inequality is used,

$$\sum_{i\in[K]} w_i^* \big[\boldsymbol{f}(\boldsymbol{w}^*, \boldsymbol{\nu}^*)\big]_i^2 - \Big(\sum_{i\in[K]} w_i^* \big[\boldsymbol{f}(\boldsymbol{w}^*, \boldsymbol{\nu}^*)\big]_i \Big)^2 = \sum_{i,j} w_i^* w_j^* \Big(\big[\boldsymbol{f}(\boldsymbol{w}^*, \boldsymbol{\nu}^*)\big]_i - \big[\boldsymbol{f}(\boldsymbol{w}^*, \boldsymbol{\nu}^*)\big]_j \Big)^2.$$

Vanishing of this quantity forces the bracketed components to be equal on $J = \mathrm{Supp}(\boldsymbol{w}^*)$. By Euler's identity $\sum_i w_i^* [\boldsymbol{f}(\boldsymbol{w}^*, \boldsymbol{\nu}^*)]_i = F(\boldsymbol{w}^*, \boldsymbol{\nu}^*)$, that common value is $F(\boldsymbol{w}^*, \boldsymbol{\nu}^*)$, yielding the equality condition (40).

For any $\boldsymbol{w} \in \Delta_J$, concavity of $F(\cdot, \boldsymbol{\nu}^*)$ gives

$$F(\boldsymbol{w}, \boldsymbol{\nu}^*) \ \leq \ F(\boldsymbol{w}^*, \boldsymbol{\nu}^*) + \boldsymbol{f}(\boldsymbol{w}^*, \boldsymbol{\mu}^*)^\top (\boldsymbol{w} - \boldsymbol{w}^*) = F(\boldsymbol{w}^*, \boldsymbol{\nu}^*),$$

since the gradient components are equal on J and $\sum_{i\in J}(w_i - w_i^*) = 0$. Because $\min_x D_x^w(\boldsymbol{w}) \leq F(\boldsymbol{w}, \boldsymbol{\nu}^*)$ for all $\boldsymbol{w}$, with equality at $\boldsymbol{w}^*$ (as $\boldsymbol{\nu}^* \in \mathtt{det}(\boldsymbol{w}^*)$), we obtain

$$\min_x D_x^w(\boldsymbol{w}) \leq \min_x D_x^w(\boldsymbol{w}^*), \quad \forall \boldsymbol{w} \in \Delta_J,$$

proving the facewise maximization in (ii).

The description and finiteness of stationary values then follow by enumerating nonempty faces $J \subseteq [K]$. □

Remark 2 (Weak Sard property)**.** *Let $g(\boldsymbol{w}) := \min_{x\in\mathcal{X}} D_x(\boldsymbol{w})$. A standard first-order (Clarke) stationarity condition for the constrained maximization $\max_{\boldsymbol{w}\in\Delta_K} g(\boldsymbol{w})$ is the normal-cone inclusion*

$$\mathbf{0} \in -\partial^\circ g(\boldsymbol{w}^*) + N_{\Delta_K}(\boldsymbol{w}^*), \tag{41}$$

where for any closed convex $C \subseteq \mathbb{R}^K$ and $\boldsymbol{w} \in C$ the normal cone is

$$N_C(\boldsymbol{w}) := \{\boldsymbol{u} \in \mathbb{R}^K : \langle \boldsymbol{u}, \boldsymbol{q} - \boldsymbol{w}\rangle \le 0 \ \ \forall \boldsymbol{q} \in C\}.$$

For $C = \Delta_K$ and $J = \mathrm{Supp}(\boldsymbol{w}^*)$, one has

$$N_{\Delta_K}(\boldsymbol{w}^*) = \Big\{\boldsymbol{u} : \exists \lambda \in \mathbb{R} \text{ s.t. } u_i = \lambda \ (i \in J),\ u_i \le \lambda \ (i \notin J)\Big\}.$$

Thus (41) *requires a Clarke subgradient to be constant on the support and no larger off the support;* (40) *in Corollary 1 recovers the equality part.*

The corresponding critical values *are the objective values attained at such points. Corollary 1 shows that this set is finite (hence has empty interior), which is exactly the "weak Sard property" postulated in Davis et al. (2020, Assumption B); see also Ioffe (2017, Section 8.4) and Bolte and Pauwels (2021, Theorem 5).*

E.2 Stochastic Approximation

This subsection casts the discrete cost-allocation recursion (11) as a stochastic approximation to the limiting differential inclusion (DI) (18). We introduce the standard continuous-time interpolation $\hat{\boldsymbol{w}}(\cdot)$ based on the random step sizes and show that, after a sufficiently large time shift, every length-T segment of the interpolation is an ε-perturbed DI solution in the sense of Benaïm's perturbation scheme (Proposition 4). As a consequence, $\hat{\boldsymbol{w}}$ is almost surely an asymptotic pseudo-trajectory (APT) of the DI, and its ω-limit set is internally chain transitive (Proposition 5). Combining this dynamical-systems structure with the Lyapunov analysis from Section E.1 (in particular, the weak Sard property of stationary values), we conclude that the Lyapunov value $\min_{x\in\mathcal{X}} D_x^w(\boldsymbol{w})$ is constant on the limit set and therefore converges along the discrete iterates (Proposition 6).

Definition 2 (Continuous-time interpolation)**.** *Set $t_0 = 0$ and $t_{n+1} = t_n + \alpha_{n+1}$ for $n \ge 0$. For $s \in [0, \alpha_{n+1}]$, define the linear interpolation of $\boldsymbol{w}_n$ by*

$$\hat{\boldsymbol{w}}(t_n + s) \triangleq \boldsymbol{w}_n + s\frac{\boldsymbol{w}_{n+1} - \boldsymbol{w}_n}{t_{n+1} - t_n}. \tag{42}$$

Then $\hat{\boldsymbol{w}}(t_n) = \boldsymbol{w}_n$ for all n.

Recall $G : \Delta_K \times \mathbb{R}^d \rightrightarrows T_{\Delta_K}$ denote the right-hand side correspondence of the continuous dynamics with the state $\boldsymbol{\theta}$:

$$G(\boldsymbol{w};\boldsymbol{\theta}) \triangleq \texttt{sel}(\boldsymbol{w};\boldsymbol{\theta}) - \boldsymbol{w},$$

where $\texttt{sel}(\boldsymbol{w};\boldsymbol{\theta})$ is defined by replacing $D_x^w(\boldsymbol{w})$ with $D_x^w(\boldsymbol{w};\boldsymbol{\theta})$. Let $d(\cdot,\cdot)$ denote the distance induced by the l_1 norm.

Definition 3 (ε-Perturbed DI)**.** *For $\varepsilon \geq 0$ define*

$$G^{\varepsilon}(\boldsymbol{w};\boldsymbol{\theta}) \triangleq \Big\{\boldsymbol{y} \in \mathbb{R}^K : \exists\ \boldsymbol{w}' \in \Delta_K \ \textit{s.t.}\ \|\boldsymbol{w} - \boldsymbol{w}'\|_1 + d(\boldsymbol{y}, G(\boldsymbol{w}';\boldsymbol{\theta})) \leq \varepsilon\Big\}.$$

Define the solution set $\mathcal{T}_{\boldsymbol{\theta}}^{\varepsilon,T} : \Delta_K \rightrightarrows \mathcal{AC}([0,T];\mathbb{R}^K)$ that maps an initial condition in Δ_K to the (nonempty) set of absolutely continuous solutions of the ε-perturbed DI on $[0,T]$:

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{w}(t) \in G^{\varepsilon}(\boldsymbol{w}(t);\boldsymbol{\theta}) + U(t), \qquad \boldsymbol{w}(0) = \boldsymbol{w}_0,$$

where locally integrable (stochastic) process $U(t)$ satisfying

$$\sup_{0\leq t\leq T}\left\|\int_0^t U(s)\ \mathrm{d}s\right\|_1 \leq \epsilon, \ \textit{with probability one.}$$

To show that our allocation update indeed realizes a perturbed differential inclusion in the sense of Definition 3, we first use the following consistency property of the instance estimator.

Lemma 9 (Sufficient exploration ensures consistency of $\boldsymbol{\theta}_n$)**.** *Under the forced-exploration schedule in* (7)*, the instance estimate satisfies $\boldsymbol{\theta}_n \to \boldsymbol{\theta}$ almost surely.*

Assuumption 3 implies that there exists a neighborhood U of the true $\boldsymbol{\theta}$ such that for all $\boldsymbol{\vartheta} \in U$, we have $\mathcal{X}(\boldsymbol{\vartheta}) = \mathcal{X}(\boldsymbol{\theta})$ and $\mathrm{Alt}_x(\boldsymbol{\vartheta}) = \mathrm{Alt}_x(\boldsymbol{\theta})$ for all x in $\mathcal{X}(\boldsymbol{\theta})$. In view of Lemma 1, the minimizer is unique for $\boldsymbol{w} \in \mathcal{R}$. Then, Degenne and Koolen (2019, Theorem 4) implies that, for a fixed true parameter $\boldsymbol{\theta} \in \mathbb{R}^d$ and for every $\boldsymbol{w} \in \mathcal{R}$ and every $x \in \mathcal{X}$, the maps

$$\boldsymbol{\vartheta} \ \mapsto\ D_x^w(\boldsymbol{w},\boldsymbol{\vartheta}) \quad \text{and} \quad \boldsymbol{\vartheta} \ \mapsto\ \nabla_{\boldsymbol{w}} D_x^w(\boldsymbol{w},\boldsymbol{\vartheta})$$

are continuous at $\boldsymbol{\vartheta} = \boldsymbol{\theta}$.

Lemma 10 (U.h.c. at the true state parameter)**.** *Fix $\boldsymbol{w} \in \mathcal{R}$ (so $\min_{x\in\mathcal{X}} D_x^w(\boldsymbol{w};\boldsymbol{\theta}) > 0$). Then $\boldsymbol{\theta}' \mapsto G(\boldsymbol{w},\boldsymbol{\theta}')$ is upper hemicontinuous at $\boldsymbol{\theta}$: for every $\varepsilon > 0$ there exists $\delta > 0$ such that*

$$\|\boldsymbol{\theta}' - \boldsymbol{\theta}\|_\infty \leq \delta \implies G(\boldsymbol{w},\boldsymbol{\theta}') \subseteq G(\boldsymbol{w};\boldsymbol{\theta}) + \varepsilon B,$$

where B is the unit ball of $\ell_1(\mathbb{R}^K)$.

Proof. Recall that $F(\boldsymbol{w},\boldsymbol{\nu};\boldsymbol{\vartheta}) = \sum_x \nu_x D_x^w(\boldsymbol{w};\boldsymbol{\vartheta})$ is continuous in $(\boldsymbol{\nu};\boldsymbol{\vartheta})$ and linear in $\boldsymbol{\nu}$ on the compact set $\Delta_{|\mathcal{X}|}$; hence by Berge's maximum theorem

$$\texttt{det}(\boldsymbol{w};\boldsymbol{\vartheta}) \triangleq \operatorname*{arg\,min}_{\boldsymbol{\nu}\in\Delta_{|\mathcal{X}|}} F(\boldsymbol{w},\boldsymbol{\nu};\boldsymbol{\vartheta})$$

is nonempty, compact-valued, and u.h.c. at $\boldsymbol{\theta}$. Since $\min_x D_x^w(\boldsymbol{w};\boldsymbol{\theta}) > 0$ and $D_x^w(\boldsymbol{w},\cdot)$ is continuous at $\boldsymbol{\theta}$, there exists a neighborhood U of $\boldsymbol{\theta}$ with $\sum_x \nu_x D_x^w(\boldsymbol{w};\boldsymbol{\vartheta}) \geq \min_{x\in\mathcal{X}} D_x^w(\boldsymbol{w};\boldsymbol{\theta})/2 > 0$ for all $\vartheta \in U$ and all $\boldsymbol{\nu} \in \mathtt{det}(\boldsymbol{w};\boldsymbol{\vartheta})$. Thus

$$(\boldsymbol{\nu};\boldsymbol{\vartheta}) \mapsto \boldsymbol{h}(\boldsymbol{w},\boldsymbol{\nu};\boldsymbol{\vartheta}) = \left(\frac{w_i \sum_x \nu_x [\nabla_{\boldsymbol{w}} D_x^w(\boldsymbol{w};\boldsymbol{\vartheta})]_i}{\sum_x \nu_x D_x^w(\boldsymbol{w};\boldsymbol{\vartheta})}\right)_{i\in[K]}$$

is continuous on $\Delta_{|\mathcal{X}|} \times U$. The image correspondence

$$\vartheta \;\mapsto\; \{\boldsymbol{h}(\boldsymbol{w},\boldsymbol{\nu};\boldsymbol{\vartheta}) - \boldsymbol{w} : \boldsymbol{\nu} \in \mathtt{det}(\boldsymbol{w};\boldsymbol{\vartheta})\} = G(\boldsymbol{w};\boldsymbol{\vartheta})$$

is therefore the continuous image of a compact-valued u.h.c. correspondence, hence u.h.c. at $\boldsymbol{\theta}$. The stated ε-inclusion follows. □

Define

$$\mathsf{NF}_n \triangleq \mathbb{1}\left\{\sqrt{\lfloor n/K \rfloor} \notin \mathbb{Z}\right\} \quad \text{and} \quad \mathsf{F}_n \triangleq \mathbb{1}\left\{\sqrt{\lfloor n/K \rfloor} \in \mathbb{Z}\right\},$$

i.e., $\mathsf{NF}_n = 1$ on non-forced rounds and $\mathsf{NF}_n = 0$ on forced rounds. Recall that $\widetilde{\boldsymbol{h}}_n \in \Delta_K$ is the cost-weighted sampling distribution defined by $\mathbb{E}[\alpha_{n+1}\boldsymbol{e}_{I_{n+1}} \mid \mathcal{H}_n] = \widetilde{\alpha}_{n+1}\widetilde{\boldsymbol{h}}_n$ with $\widetilde{\alpha}_{n+1} = \mathbb{E}[\alpha_{n+1} \mid \mathcal{H}_n]$.

Lemma 11 (Martingale noise on non-forced rounds)**.** *Define*

$$\boldsymbol{\xi}_m \triangleq \sum_{n=0}^{m-1} \mathsf{NF}_n \alpha_{n+1}\Big(\boldsymbol{e}_{I_{n+1}} - \widetilde{\boldsymbol{h}}_n\Big), \qquad m \geq 0.$$

Then $(\boldsymbol{\xi}_m)_{m\geq 0}$ is an $\mathbb{R}^K$-valued martingale with respect to $(\mathcal{H}_m)_{m\geq 0}$, and $\boldsymbol{\xi}_m$ converges almost surely (and in L^2) to a finite random limit $\boldsymbol{\xi}_\infty$.

Proof. For each $n \geq 0$,

$$\mathbb{E}\left[\alpha_{n+1}\boldsymbol{e}_{I_{n+1}} \mid \mathcal{H}_n\right] = \widetilde{\alpha}_{n+1}\widetilde{\boldsymbol{h}}_n \quad \text{and} \quad \mathbb{E}[\alpha_{n+1} \mid \mathcal{H}_n] = \widetilde{\alpha}_{n+1}.$$

Hence

$$\mathbb{E}\Big[\mathsf{NF}_n \alpha_{n+1}(\boldsymbol{e}_{I_{n+1}} - \widetilde{\boldsymbol{h}}_n) \mid \mathcal{H}_n\Big] = \mathsf{NF}_n\Big(\widetilde{\alpha}_{n+1}\widetilde{\boldsymbol{h}}_n - \widetilde{\alpha}_{n+1}\widetilde{\boldsymbol{h}}_n\Big) = \mathbf{0},$$

so $(\boldsymbol{\xi}_m)$ is a martingale.

Moreover, $\|\boldsymbol{e}_{I_{n+1}} - \widetilde{\boldsymbol{h}}_n\|_1 \leq 2$, so

$$\|\mathsf{NF}_n \alpha_{n+1}(\boldsymbol{e}_{I_{n+1}} - \widetilde{\boldsymbol{h}}_n)\|_1^2 \leq 4\alpha_{n+1}^2.$$

Using $\alpha_{n+1} \leq \frac{c_{\max}}{c_{\min}(n+1)}$ (from Assumption 2),

$$\sum_{n=0}^{\infty} \mathbb{E}\left[\|\mathsf{NF}_n \alpha_{n+1}(\boldsymbol{e}_{I_{n+1}} - \widetilde{\boldsymbol{h}}_n)\|_1^2\right] \leq 4\sum_{n=0}^{\infty} \alpha_{n+1}^2 \leq 4\Big(\frac{c_{\max}}{c_{\min}}\Big)^2 \sum_{n=0}^{\infty} \frac{1}{(n+1)^2} < \infty.$$

Thus each coordinate martingale has square-summable increments, and by the martingale convergence theorem, $\boldsymbol{\xi}_m \to \boldsymbol{\xi}_\infty$ almost surely (and in L^2) for some finite random vector $\boldsymbol{\xi}_\infty$. □

Proposition 4 (Interpolation process is a perturbed solution)**.** *Fix $\varepsilon > 0$ and $T > 0$. Then there exists an almost surely finite $S(\varepsilon, T)$ such that for all $s \geq S(\varepsilon, T)$, the translated interpolation segment $t \mapsto \hat{\boldsymbol{w}}(s+t)$ on $[0,T]$ belongs to $\mathcal{T}_{\boldsymbol{\theta}}^{\varepsilon,T}(\hat{\boldsymbol{w}}(s))$. Equivalently, for all such s there exists a locally integrable process $U_s : [0,T] \to \mathbb{R}^K$ satisfying*

$$\frac{d}{dt}\hat{\boldsymbol{w}}(s+t) \in G^{\varepsilon}\big(\hat{\boldsymbol{w}}(s+t); \boldsymbol{\theta}\big) + U_s(t) \quad \text{for a.e. } t \in [0,T],$$

and

$$\sup_{0 \leq u \leq T} \left\| \int_0^u U_s(r) dr \right\|_1 \leq \varepsilon.$$

Proof. For notational convenience, define x_n by the detection rule $x_n \in \arg\min_{x \in \mathcal{X}(\boldsymbol{\theta}_n)} D_x^w(\boldsymbol{w}_n; \boldsymbol{\theta}_n)$ for every n (even on forced rounds); this is $\mathcal{H}_n$-measurable.

Let $t \in (t_n, t_{n+1})$. By the definition of the interpolation and $t_{n+1} - t_n = \alpha_{n+1}$,

$$\dot{\hat{\boldsymbol{w}}}(t) = \frac{\boldsymbol{w}_{n+1} - \boldsymbol{w}_n}{\alpha_{n+1}} = \boldsymbol{e}_{I_{n+1}} - \boldsymbol{w}_n.$$

Define the (deterministic) drift proxy and noise on (t_n, t_{n+1}) by

$$\boldsymbol{v}(t) \triangleq \boldsymbol{h}^{x_n}(\boldsymbol{w}_n; \boldsymbol{\theta}_n) - \boldsymbol{w}_n, \qquad \boldsymbol{U}(t) \triangleq \boldsymbol{e}_{I_{n+1}} - \boldsymbol{h}^{x_n}(\boldsymbol{w}_n; \boldsymbol{\theta}_n),$$

so that $\dot{\hat{\boldsymbol{w}}}(t) = \boldsymbol{v}(t) + \boldsymbol{U}(t)$.

Step 1: $\boldsymbol{v}(t) \in G^{\varepsilon}(\hat{\boldsymbol{w}}(t); \boldsymbol{\theta})$ for all large n. For $n \geq K$ we have $\boldsymbol{w}_n \in \mathrm{int}(\Delta_K) \subseteq \mathcal{R}$, so $\boldsymbol{h}^{x_n}(\boldsymbol{w}_n; \boldsymbol{\theta}_n)$ is well defined. Since x_n is a minimizer of $D_x^w(\boldsymbol{w}_n; \boldsymbol{\theta}_n)$, the point mass $\boldsymbol{\nu} = e_{x_n}$ belongs to $\mathtt{det}(\boldsymbol{w}_n; \boldsymbol{\theta}_n)$, and hence $\boldsymbol{h}^{x_n}(\boldsymbol{w}_n; \boldsymbol{\theta}_n) \in \mathtt{sel}(\boldsymbol{w}_n; \boldsymbol{\theta}_n)$. Therefore $\boldsymbol{v}(t) \in G(\boldsymbol{w}_n; \boldsymbol{\theta}_n)$.

Also, for any $r \in [0, \alpha_{n+1}]$,

$$\|\hat{\boldsymbol{w}}(t_n + r) - \boldsymbol{w}_n\|_1 = \frac{r}{\alpha_{n+1}} \|\boldsymbol{w}_{n+1} - \boldsymbol{w}_n\|_1 = \frac{r}{\alpha_{n+1}} \alpha_{n+1} \|\boldsymbol{e}_{I_{n+1}} - \boldsymbol{w}_n\|_1 \leq 2\alpha_{n+1}.$$

By Lemma 9, $\boldsymbol{\theta}_n \to \boldsymbol{\theta}$ a.s., and by Lemma 10, for every $\eta > 0$ there exists (a.s. finite) $N_1(\eta)$ such that for all $n \geq N_1(\eta)$,

$$G(\boldsymbol{w}_n; \boldsymbol{\theta}_n) \subseteq G(\boldsymbol{w}_n; \boldsymbol{\theta}) + \eta B,$$

where B is the ℓ_1 unit ball. Hence for all $n \geq N_1(\eta)$ and all $t \in (t_n, t_{n+1})$,

$$d\big(\boldsymbol{v}(t), G(\boldsymbol{w}_n; \boldsymbol{\theta})\big) \leq \eta.$$

Choose $\eta = \varepsilon/2$, and then choose $N_2(\varepsilon)$ such that $2\alpha_{n+1} \leq \varepsilon/2$ for all $n \geq N_2(\varepsilon)$. For $n \geq N_0 := \max\{N_1(\varepsilon/2), N_2(\varepsilon)\}$ and all $t \in (t_n, t_{n+1})$, taking $\boldsymbol{w}' = \boldsymbol{w}_n$ in the definition of G^{ε} gives

$$\|\hat{\boldsymbol{w}}(t) - \boldsymbol{w}'\|_1 + d\big(\boldsymbol{v}(t), G(\boldsymbol{w}'; \boldsymbol{\theta})\big) \leq \varepsilon/2 + \varepsilon/2 = \varepsilon,$$

i.e., $\boldsymbol{v}(t) \in G^{\varepsilon}\big(\hat{\boldsymbol{w}}(t); \boldsymbol{\theta}\big)$.

Step 2: uniform smallness of the integrated noise. Fix $s \geq t_{N_0}$ and define the shift $U_s(t) := U(s+t)$ on $[0,T]$. We will show

$$\sup_{0\leq u\leq T}\left\|\int_0^u U_s(r)dr\right\|_1 = \sup_{0\leq u\leq T}\left\|\int_s^{s+u} U(r)dr\right\|_1 \leq \varepsilon$$

for all sufficiently large s.

Decompose, for $t \in (t_n, t_{n+1})$,

$$\boldsymbol{U}(t) = \underbrace{(\boldsymbol{e}_{I_{n+1}} - \widetilde{\boldsymbol{h}}_n)}_{\boldsymbol{M}(t)} + \underbrace{(\widetilde{\boldsymbol{h}}_n - \boldsymbol{h}^{x_n}(\boldsymbol{w}_n;\boldsymbol{\theta}_n))}_{\boldsymbol{B}(t)}.$$

Then $\boldsymbol{M}(t)$ is constant on (t_n, t_{n+1}) and

$$\int_{t_n}^{t_{n+1}} \boldsymbol{M}(t)dt = \alpha_{n+1}(\boldsymbol{e}_{I_{n+1}} - \widetilde{\boldsymbol{h}}_n).$$

Thus the integrated martingale term over any union of whole intervals is a difference of the martingale $\boldsymbol{\xi}_m$ from Lemma 11. Since $\boldsymbol{\xi}_m \to \boldsymbol{\xi}_\infty$ a.s., we have

$$\Delta_\xi(n) := \sup_{m\geq n}\|\boldsymbol{\xi}_m - \boldsymbol{\xi}_n\|_1 \xrightarrow[n\to\infty]{} 0 \quad \text{a.s.}$$

Next, we claim that $\sum_{n=0}^\infty \alpha_{n+1}\|\widetilde{\boldsymbol{h}}_n - \boldsymbol{h}^{x_n}(\boldsymbol{w}_n;\boldsymbol{\theta}_n)\|_1 < \infty$ a.s. Indeed, on non-forced rounds, Lemma 2 gives $\|\widetilde{\boldsymbol{h}}_n - \boldsymbol{h}^{x_n}(\boldsymbol{w}_n;\boldsymbol{\theta}_n)\|_1 \leq \frac{c_{\max}}{nc_{\min}}$, so using $\alpha_{n+1} \leq \frac{c_{\max}}{c_{\min}(n+1)}$ we obtain a summable bound of order $1/n^2$. On forced rounds, $\|\widetilde{\boldsymbol{h}}_n - \boldsymbol{h}^{x_n}(\boldsymbol{w}_n;\boldsymbol{\theta}_n)\|_1 \leq 2$, and the forced schedule satisfies

$$\sum_{n:\mathsf{F}_n=1}\alpha_{n+1} \;\leq\; \frac{c_{\max}}{c_{\min}}\sum_{m=1}^{\infty}\sum_{k=0}^{K-1}\frac{1}{Km^2+k+1} \;<\; \infty.$$

Therefore the bias series is absolutely summable, and its tail

$$\Delta_B(n) := \sum_{k=n}^{\infty}\alpha_{k+1}\|\widetilde{\boldsymbol{h}}_k - \boldsymbol{h}^{x_k}(\boldsymbol{w}_k;\boldsymbol{\theta}_k)\|_1$$

satisfies $\Delta_B(n) \to 0$ a.s.

Finally, the contribution from the two boundary (partial) intervals at the start and end of $[s, s+u]$ is bounded by $4\sup_{k\geq n(s)}\alpha_{k+1}$ since $\|U(t)\|_1 \leq 2$. Because $\alpha_{n+1} \to 0$, this boundary term vanishes as $s \to \infty$.

Putting these together: let $n(s) := \max\{n : \; t_n \leq s\}$. Then for any $u \in [0,T]$,

$$\left\|\int_s^{s+u} U(t)dt\right\|_1 \;\leq\; 4\sup_{k\geq n(s)}\alpha_{k+1} \;+\; \Delta_\xi(n(s)) \;+\; \Delta_B(n(s)).$$

Each of the three terms tends to 0 a.s. as $s \to \infty$. Hence there exists an a.s. finite $S(\varepsilon, T)$ such that for all $s \geq S(\varepsilon, T)$ the right-hand side is at most ε, proving the desired noise bound.

Consequently, for all $s \geq S(\varepsilon, T)$, we have $\dot{\hat{\boldsymbol{w}}}(s+t) = \boldsymbol{v}(s+t) + U_s(t)$ with $\boldsymbol{v}(s+t) \in G^\varepsilon(\hat{\boldsymbol{w}}(s+t);\boldsymbol{\theta})$ for a.e. $t \in [0,T]$ and $\sup_{0\leq u\leq T}\|\int_0^u U_s(r)dr\|_1 \leq \varepsilon$. Thus $\hat{\boldsymbol{w}}(s+\cdot) \in \mathcal{T}_{\boldsymbol{\theta}}^{\varepsilon,T}(\hat{\boldsymbol{w}}(s))$. □

Definition 4 (Asymptotic pseudo-trajectory (APT) for u.h.c DI)**.** *Consider a u.h.c. differential inclusion and let $S_{x(0)}$ be the set of solutions with initial value $x(0)$; write $S \triangleq \bigcup_{x(0)} S_{x(0)}$ for the set of all solutions. A bounded continuous curve $z : [0, \infty) \to \mathbb{R}^m$ is an* asymptotic pseudo-trajectory (APT) *of the DI if for every $T > 0$,*

$$\lim_{t\to\infty} \inf_{\sigma\in S} \sup_{s\in[0,T]} \|z(t+s) - \sigma(s)\| = 0.$$

Equivalently: for all $\varepsilon, T > 0$ there exists $t_0 < \infty$ such that for all $t \geq t_0$ there is a DI solution segment $\sigma : [0, T] \to \mathbb{R}^m$ with $\sup_{s\in[0,T]} \|z(t+s) - \sigma(s)\| \leq \varepsilon$.

Note that in this definition the comparison solution σ need not start exactly from $z(t)$; we only require that its initial condition be within ε of $z(t)$. This is precisely the definition adopted in Esponda et al. (2022), and see also Appendix C of Bianchi et al. (2024).

Definition 5 ((ε, T)-chains and internally chain transitive (ICT) sets)**.** *Let (X, d) be a compact metric space and $\{\Phi_t\}_{t\geq 0}$ a continuous semiflow on X. For $\varepsilon > 0$ and $T > 0$, an (ε, T)-chain from x to y is a finite sequence*

$$x = z_0,\ z_1, \ldots, z_m = y, \qquad t_1, \ldots, t_m \geq T,$$

such that $d\big(\Phi_{t_i}(z_{i-1}), z_i\big) < \varepsilon$ for all i. A nonempty compact set $L \subset X$ is ICT *if: (i) L is invariant ($\Phi_t(L) = L$ for all $t \geq 0$); and (ii) for all $x, y \in L$ and all $\varepsilon, T > 0$, there exists an (ε, T)-chain entirely contained in L from x to y.*

Proposition 5 (Interpolation ⇒ APT ⇒ ICT ω-limit set)**.** *The interpolation process $\hat{\boldsymbol{w}}$ is an a.s. asymptotic pseudo-trajectory of the differential inclusion* (18)*; and its ω-limit set $\Omega \triangleq \omega(\hat{\boldsymbol{w}}) \subset \Delta_K$ is internally chain transitive.*

Proof. The first part follows from our Proposition 4 and Benaïm et al. (2005, Theorem 4.2). The second conclusion follows from Benaïm et al. (2005, Theorem 4.3). □

Proposition 6 (Constancy of $\min_x D_x$ on ICT sets)**.** *The map $\boldsymbol{w} \mapsto \min_{x\in\mathcal{X}} D_x^w(\boldsymbol{w})$ is constant on Ω. Consequently (along the* `PAN` *iterates), $\min_{x\in\mathcal{X}} D_x^w(\boldsymbol{w}_n; \boldsymbol{\theta}_n)$ converges almost surely.*

Proof. By Theorem 4, $t \mapsto \min_x D_x^w(\boldsymbol{w}(t))$ is nondecreasing along any solution and strictly increasing off the stationary set of Corollary 1; the set of stationary values has empty interior. Hence Proposition 3.27 of Benaïm et al. (2005) applies: on the ICT set $\Omega = \omega(\hat{\boldsymbol{w}})$ the function $\min_x D_x^w$ is constant. □

Remark 3. *The framework of Benaïm et al. (2005) using the* internally chain transitive *set is essentially the same as the "non-escape argument" in Davis et al. (2020, Section 3.3). In Benaïm's framework, an ICT set is a compact invariant set with no proper attracting subset, see Benaïm et al. (2005, Proposition 3.20). If a Lyapunov function is nondecreasing along the DI and its stationary values form a thin set (weak Sard), then every small sublevel that is forward invariant would have to attract the whole ICT set; this forces the Lyapunov to be constant on that set. The proof in Davis et al. (2020, Section 3.3) does the similar thing: they show the APT cannot keep re-entering them—hence "non-escape"—so the Lyapunov must converge and be constant on the limit set.*

E.3 Algorithm 1 Attains Positive Information

The stochastic-approximation analysis in Section E.2 implies that the Lyapunov value converges almost surely to a stationary value of the limiting differential inclusion. A priori, this limit could be the degenerate value 0, corresponding to vanishing discriminative information as the allocation approaches the simplex boundary. In this subsection we rule out this pathology. Using the gradient lower bound from Lemma 1 and the asymptotic accuracy of the IDS sampling rule, we show that whenever the information value becomes too small, the induced sampling dynamics creates a uniform multiplicative "lift" on the geometric mean $G_n = \prod_{i=1}^K w_{n,i}$ on non-forced rounds, while forced rounds can only decrease G_n by a summable amount. Since G_n is uniformly bounded above on the simplex, this yields a contradiction, and therefore the limiting information value must be bounded away from 0 by an explicit instance-dependent constant.

Recall from Lemma 1 that for each $x \in \mathcal{X}$ there exists $d_x(\boldsymbol{\theta}) > 0$ such that $\sum_{i=1}^K [\nabla_{\boldsymbol{w}} D_x^w(\boldsymbol{w};\boldsymbol{\theta})]_i \geq d_x(\boldsymbol{\theta})$ on $\mathrm{int}(\Delta_K)$. We set

$$d_* \triangleq \min_{x\in\mathcal{X}} d_x(\boldsymbol{\theta}) \;>\; 0. \tag{43}$$

Proposition 7 (Interior start rules out zero information)**.** *There exists an instance-dependent constant $D_* = D_*(\boldsymbol{\theta})$ such that*

$$\lim_{n\to\infty} \; D_{x_n}^w(\boldsymbol{w}_n;\boldsymbol{\theta}_n) \;\geq\; D_* \triangleq \frac{c_{\min}^2 d_*}{8c_{\max}^2(K+2)} \quad \textit{with probability one}$$

where the limit exists by Proposition 6 and Lemma 2.

Proof. We consider only $n \geq K$ so that $\boldsymbol{w}_n \in \mathrm{int}(\Delta_K)$. Assume for contradiction that on a set A with $\mathbb{P}(A) > 0$, $\lim_{n\to\infty} \min_x D_x^w(\boldsymbol{w}_n;\boldsymbol{\theta}_n) < D_*$.

Non-forced rounds give a uniform multiplicative lift. Conditioning on $\mathcal{H}_n$ and for non-forced rounds n large enough such that $\widetilde{h}_{n,i} \geq h_i^{x_n}(\boldsymbol{w}_n;\boldsymbol{\theta}_n)/2 = w_{n,i}[\nabla_{\boldsymbol{w}} D_{x_n}^w(\boldsymbol{w}_n;\boldsymbol{\theta}_n)]_i / 2D_{x_n}^w(\boldsymbol{w}_n;\boldsymbol{\theta}_n)$,

$$\begin{aligned}
\mathbb{E}\left[\sum_{i=1}^K \frac{w_{n+1,i}}{w_{n,i}} \;\middle|\; \mathcal{H}_n\right] &= K + \widetilde{\alpha}_{n+1}\left(\sum_{i=1}^K \frac{\widetilde{h}_{n,i}}{w_{n,i}} - K\right) \\
&\geq K + \widetilde{\alpha}_{n+1}\left(\frac{\sum_i [\nabla_{\boldsymbol{w}} D_{x_n}^w(\boldsymbol{w}_n;\boldsymbol{\theta}_n)]_i}{2D_{x_n}^w(\boldsymbol{w}_n;\boldsymbol{\theta}_n)} - K\right).
\end{aligned}$$

By (43), continuity in $\boldsymbol{\theta}$, and $\boldsymbol{\theta}_n \to \boldsymbol{\theta}$ a.s., there exists n_0 such that, on A and for all non-forced $n \geq n_0$,

$$\sum_i [\nabla_{\boldsymbol{w}} D_{x_n}^w(\boldsymbol{w}_n;\boldsymbol{\theta}_n)]_i \geq d_*/2, \qquad D_{x_n}^w(\boldsymbol{w}_n;\boldsymbol{\theta}_n) \leq 2D_* = \frac{c_{\min}^2 d_*}{4c_{\max}^2(K+2)}.$$

Hence, for non-forced $n \geq n_0$,

$$\mathbb{E}\left[\sum_{i=1}^K \frac{w_{n+1,i}}{w_{n,i}} \;\middle|\; \mathcal{H}_n\right] \geq K + \frac{2c_{\max}^2}{c_{\min}^2}\widetilde{\alpha}_{n+1} \geq K + 2\frac{c_{\max}}{c_{\min}(n+1)}. \tag{44}$$

Since all the ratios satisfy $w_{n+1,i}/w_{n,i} \geq 1 - \frac{c_{\max}}{c_{\min}(n+1)}$, by the Bauer maximum principle (concavity of $\sum_i \log(\cdot)$) with $G_n \triangleq \prod_{i=1}^K w_{n,i}$ we have

$$\frac{G_{n+1}}{G_n} = \prod_{i=1}^K \frac{w_{n+1,i}}{w_{n,i}} \geq \left(1 - \frac{c_{\max}}{c_{\min}(n+1)}\right)^{K-1} \left(\sum_{i=1}^K \frac{w_{n+1,i}}{w_{n,i}} - (K-1)\left(1 - \frac{c_{\max}}{c_{\min}(n+1)}\right)\right).$$

Since this inequality holds for every realization conditional on $\mathcal{H}_n$, we may take conditional expectations on both sides and use (44) for the sum to obtain, for all large non-forced n,

$$\mathbb{E}\left[\frac{G_{n+1}}{G_n} \;\middle|\; \mathcal{H}_n\right] \geq \left(1 - \frac{c_{\max}}{c_{\min}(n+1)}\right)^{K-1}\left(1 + \frac{(K+1)c_{\max}}{(n+1)c_{\min}}\right) \geq 1 + \frac{c_{\max}}{c_{\min}(n+1)}, \tag{45}$$

where in the last inequality we used $(1-x)^{K-1} \geq 1-(K-1)x$ and $(1-(K-1)x)(1+(K+1)x) = 1 + 2x - (K^2-1)x^2 \geq 1 + x$ for $x = \frac{c_{\max}}{c_{\min}(n+1)}$ and n large enough.

Forced rounds only decrease mildly. At a forced round, since $w_{n+1,i} \geq (1 - c_{\max}/(n+1)c_{\min})w_{n,i}$, we have the deterministic bound

$$\mathbb{E}\left[\frac{G_{n+1}}{G_n} \;\middle|\; \mathcal{H}_n\right] = \mathbb{E}\left[\prod_{i=1}^K \frac{w_{n+1,i}}{w_{n,i}} \;\middle|\; \mathcal{H}_n\right] \geq \left(1 - \frac{c_{\max}}{c_{\min}(n+1)}\right)^K \geq 1 - \frac{c_{\max}K}{c_{\min}(n+1)}.$$

Recall the event

$$A \triangleq \Big\{\lim_{n\to\infty} \min_{x\in\mathcal{X}} D_x^w(\boldsymbol{w}_n; \boldsymbol{\theta}_n) < D_*\Big\},$$

and suppose for contradiction that $\mathbb{P}(A) > 0$.

On the almost sure event where $\boldsymbol{\theta}_n \to \boldsymbol{\theta}$ and Lemma 2 holds, the bounds used above (in particular, the existence of n_0 ensuring $\sum_i [\nabla_{\boldsymbol{w}} D_{x_n}^w(\boldsymbol{w}_n; \boldsymbol{\theta}_n)]_i \geq d_*/2$ and $D_{x_n}^w(\boldsymbol{w}_n; \boldsymbol{\theta}_n) \leq 2D_*$ on A for all large non-forced n, as well as $\widetilde{h}_{n,i} \geq h_i^{x_n}(\boldsymbol{w}_n; \boldsymbol{\theta}_n)/2$) imply that there exists a (random) time $N_0 < \infty$ such that, on A, the one-step inequality

$$\mathbb{E}[G_{n+1} \mid \mathcal{H}_n] \;\geq\; a_{n+1} G_n \qquad \text{for all } n \geq N_0 \tag{46}$$

holds, where a_{n+1} is defined as

$$a_{n+1} := \begin{cases} 1 + \dfrac{c_{\max}}{c_{\min}(n+1)}, & \sqrt{\left\lfloor \dfrac{n}{K} \right\rfloor} \notin \mathbb{Z}, \\ 1 - \dfrac{c_{\max}K}{c_{\min}(n+1)}, & \sqrt{\left\lfloor \dfrac{n}{K} \right\rfloor} \in \mathbb{Z}. \end{cases}$$

Moreover, for any deterministic $N \geq K$ we have $G_N = \prod_{i=1}^K w_{N,i} > 0$ almost surely.

Define the tail event

$$A_N \triangleq A \cap \{N_0 \leq N\}.$$

Since $A = \bigcup_{N\geq 1} A_N$, there exists a *deterministic* $N \geq K$ such that $\mathbb{P}(A_N) > 0$. Fix such an N for the remainder of the argument. On A_N, the inequality (46) holds for *all* $n \geq N$.

Now fix any $m > N$ and consider the conditional law given $(\mathcal{H}_N, A_N)$. Since A_N enforces (46) for every step $n \geq N$, the tower property yields, for each $n \geq N$,

$$\mathbb{E}[G_{n+1} \mid \mathcal{H}_N, A_N] = \mathbb{E}\left[\mathbb{E}[G_{n+1} \mid \mathcal{H}_n] | \mathcal{H}_N, A_N\right] \;\geq\; a_{n+1}\mathbb{E}[G_n \mid \mathcal{H}_N, A_N].$$

Iterating from $n = N$ to $m-1$ gives

$$\mathbb{E}[G_m \mid \mathcal{H}_N, A_N] \;\geq\; G_N \prod_{n=N}^{m-1} a_{n+1}.$$

Multiplying by $\mathbb{P}(A_N \mid \mathcal{H}_N)$ and taking conditional expectations gives

$$\mathbb{E}[G_m \mathbb{1}_{A_N} \mid \mathcal{H}_N] = \mathbb{P}(A_N \mid \mathcal{H}_N)\mathbb{E}[G_m \mid \mathcal{H}_N, A_N] \;\geq\; \mathbb{P}(A_N \mid \mathcal{H}_N) G_N \prod_{n=N}^{m-1} a_{n+1}.$$

Taking expectations and using that G_N is $\mathcal{H}_N$-measurable,

$$\mathbb{E}[G_m \mathbb{1}_{A_N}] \;\geq\; \Big(\prod_{n=N}^{m-1} a_{n+1}\Big)\mathbb{E}\left[\mathbb{P}(A_N \mid \mathcal{H}_N) G_N\right] = \Big(\prod_{n=N}^{m-1} a_{n+1}\Big)\mathbb{E}[G_N \mathbb{1}_{A_N}].$$

Since $\mathbb{P}(A_N) > 0$ and $G_N > 0$ almost surely, we have $\mathbb{E}[G_N \mathbb{1}_{A_N}] > 0$.

Finally, $G_m = \prod_{i=1}^{K} w_{m,i} \leq K^{-K}$ for all m, hence $\mathbb{E}[G_m \mathbb{1}_{A_N}] \leq K^{-K}$ for all m. Therefore,

$$K^{-K} \;\geq\; \mathbb{E}[G_m \mathbb{1}_{A_N}] \;\geq\; \Big(\prod_{n=N}^{m-1} a_{n+1}\Big)\mathbb{E}[G_N \mathbb{1}_{A_N}]. \tag{47}$$

It remains to note that $\prod_{n=N}^{m-1} a_{n+1} \to \infty$ as $m \to \infty$. Indeed, for all sufficiently large n, $\log(1+x) \geq x/2$ for $x = \frac{c_{\max}}{c_{\min}(n+1)}$ and $\log(1-y) \geq -2y$ for $y = \frac{c_{\max} K}{c_{\min}(n+1)}$, so

$$\log\Big(\prod_{n=N}^{m-1} a_{n+1}\Big) \;\geq\; \sum_{n=N}^{m-1} \frac{c_{\max}}{2c_{\min}(n+1)} \mathsf{NF}_n - \sum_{n=N}^{m-1} \frac{2c_{\max} K}{c_{\min}(n+1)} \mathsf{F}_n.$$

The first sum diverges because $\sum_n \mathsf{NF}_n/(n+1) = \infty$, while the second sum converges because forced rounds occur in blocks starting at $n = Km^2$ and hence $\sum_n \mathsf{F}_n/(n+1) < \infty$. Thus $\prod_{n=N}^{m-1} a_{n+1} \to \infty$, contradicting (47) since $\mathbb{E}[G_N \mathbb{1}_{A_N}] > 0$.

Therefore $\mathbb{P}(A) = 0$, i.e., $\lim_{n\to\infty} \min_x D_x^w(\boldsymbol{w}_n; \boldsymbol{\theta}_n) \geq D_*$ almost surely, and by detection consistency $\lim_{n\to\infty} D_{x_n}^w(\boldsymbol{w}_n; \boldsymbol{\theta}_n) \geq D_*$ almost surely as claimed. □

E.4 Completing the Proof: Algorithm 1 Attains Optimal Value

In this final step we show that the limiting Lyapunov value identified in Section E.2 must coincide with the global minimax value $F^* = \Gamma^*$, rather than an arbitrary stationary value of the differential inclusion. We first formalize the minimax structure by introducing the saddle-point set $\mathcal{E}$ of the concave–convex game $F(\boldsymbol{w}, \boldsymbol{\mu})$ and select a representative saddle point whose weight vector is strictly positive on every coordinate that can be positive at equilibrium (Lemma 12). Using this representative, we define a Kullback–Leibler potential $V_n = D_{\mathrm{KL}}(\boldsymbol{w}^* \| \boldsymbol{w}_n)$ on the active face and

prove a uniform negative-drift bound for V_n whenever the value gap $F^* - \min_x D_x^w(\boldsymbol{w}_n;\boldsymbol{\theta})$ stays bounded away from zero. A Robbins–Siegmund almost-supermartingale argument then forces the gap to vanish, yielding $\limsup_n \min_x D_x^w(\boldsymbol{w}_n;\boldsymbol{\theta}) = F^*$ and, combined with the value convergence from Section E.2, establishes $\min_x D_x^w(\boldsymbol{w}_n;\boldsymbol{\theta}_n) \to F^*$ almost surely.

A pair $(\boldsymbol{w}^*, \boldsymbol{\mu}^*)$ is a *saddle point* of F if

$$F(\boldsymbol{w}, \boldsymbol{\mu}^*) \;\le\; F(\boldsymbol{w}^*, \boldsymbol{\mu}^*) \;\le\; F(\boldsymbol{w}^*, \boldsymbol{\mu}) \qquad \forall \boldsymbol{w} \in \Delta_K,\ \boldsymbol{\mu} \in \Delta_{|\mathcal{X}|}.$$

The common value $F^* \triangleq F(\boldsymbol{w}^*, \boldsymbol{\mu}^*)$ is the minimax value. Let

$$\mathcal{E} \triangleq \{(\boldsymbol{w}, \boldsymbol{\mu}) \in \Delta_K \times \Delta_{|\mathcal{X}|} : (\boldsymbol{w}, \boldsymbol{\mu}) \text{ is a saddle point of } F\}.$$

For each $i \in [K]$, define the attainable i-th coordinate set

$$\mathcal{E}_{w_i} \triangleq \Big\{ w_i \in [0,1] : \exists (\boldsymbol{w}, \boldsymbol{\mu}) \in \mathcal{E} \text{ with } (\boldsymbol{w})_i = w_i \Big\},$$

and let $\mathcal{K}_0 \triangleq \{i : \mathcal{E}_{w_i} = \{0\}\}$ and $\mathcal{K}_1 \triangleq [K] \setminus \mathcal{K}_0$.

Lemma 12 (A strictly positive representative)**.** *The saddle-point set $\mathcal{E}$ is nonempty, compact, and convex. Consequently, there exists $(\boldsymbol{w}^*, \boldsymbol{\mu}^*) \in \mathcal{E}$ such that $(\boldsymbol{w}^*)_i = 0$ for all $i \in \mathcal{K}_0$ and $(\boldsymbol{w}^*)_i > 0$ for all $i \in \mathcal{K}_1$.*

Proof. The function $F(\boldsymbol{w}, \boldsymbol{\mu}) = \sum_{x \in \mathcal{X}} \mu_x D_x^w(\boldsymbol{w})$ is continuous on the compact convex set $\Delta_K \times \Delta_{|\mathcal{X}|}$, concave in $\boldsymbol{w}$ (as a nonnegative combination of concave maps D_x^w), and linear (hence convex) in $\boldsymbol{\mu}$. Therefore, Sion's minimax theorem applies and yields

$$\max_{\boldsymbol{w} \in \Delta_K} \min_{\boldsymbol{\mu} \in \Delta_{|\mathcal{X}|}} F(\boldsymbol{w}, \boldsymbol{\mu}) = \min_{\boldsymbol{\mu} \in \Delta_{|\mathcal{X}|}} \max_{\boldsymbol{w} \in \Delta_K} F(\boldsymbol{w}, \boldsymbol{\mu}) = F^*.$$

Since the max and min are attained by compactness and continuity, a saddle point exists and hence $\mathcal{E} \neq \emptyset$.

Let $(\boldsymbol{w}^n, \boldsymbol{\mu}^n) \in \mathcal{E}$ be a convergent sequence with limit $(\boldsymbol{w}, \boldsymbol{\mu}) \in \Delta_K \times \Delta_{|\mathcal{X}|}$. For every $\boldsymbol{w}' \in \Delta_K$ and $\boldsymbol{\mu}' \in \Delta_{|\mathcal{X}|}$, saddle-point inequalities give

$$F(\boldsymbol{w}', \boldsymbol{\mu}^n) \le F(\boldsymbol{w}^n, \boldsymbol{\mu}^n) \le F(\boldsymbol{w}^n, \boldsymbol{\mu}').$$

Letting $n \to \infty$ and using continuity of F yields $F(\boldsymbol{w}', \boldsymbol{\mu}) \le F(\boldsymbol{w}, \boldsymbol{\mu}) \le F(\boldsymbol{w}, \boldsymbol{\mu}')$, so $(\boldsymbol{w}, \boldsymbol{\mu}) \in \mathcal{E}$. Thus $\mathcal{E}$ is closed. Since $\mathcal{E} \subset \Delta_K \times \Delta_{|\mathcal{X}|}$ and the latter is compact, $\mathcal{E}$ is compact.

Take two saddle points $(\boldsymbol{w}^1, \boldsymbol{\mu}^1), (\boldsymbol{w}^2, \boldsymbol{\mu}^2) \in \mathcal{E}$ and $\lambda \in [0,1]$. Define $\boldsymbol{w}^\lambda = \lambda \boldsymbol{w}^1 + (1-\lambda)\boldsymbol{w}^2$ and $\boldsymbol{\mu}^\lambda = \lambda \boldsymbol{\mu}^1 + (1-\lambda)\boldsymbol{\mu}^2$. Let $F^* = F(\boldsymbol{w}^1, \boldsymbol{\mu}^1) = F(\boldsymbol{w}^2, \boldsymbol{\mu}^2)$ denote the common saddle value.

First, for any $\boldsymbol{w}' \in \Delta_K$, using linearity of $F(\boldsymbol{w}', \cdot)$ and the saddle inequalities $F(\boldsymbol{w}', \boldsymbol{\mu}^k) \le F^*$ for $k = 1, 2$, we obtain

$$F(\boldsymbol{w}', \boldsymbol{\mu}^\lambda) = \lambda F(\boldsymbol{w}', \boldsymbol{\mu}^1) + (1-\lambda) F(\boldsymbol{w}', \boldsymbol{\mu}^2) \le \lambda F^* + (1-\lambda) F^* = F^*.$$

Second, for any $\boldsymbol{\mu}' \in \Delta_{|\mathcal{X}|}$, concavity of $F(\cdot, \boldsymbol{\mu}')$ and the saddle inequalities $F(\boldsymbol{w}^k, \boldsymbol{\mu}') \ge F^*$ give

$$F(\boldsymbol{w}^\lambda, \boldsymbol{\mu}') \;\ge\; \lambda F(\boldsymbol{w}^1, \boldsymbol{\mu}') + (1-\lambda) F(\boldsymbol{w}^2, \boldsymbol{\mu}') \;\ge\; F^*.$$

In particular, taking $\boldsymbol{\mu}' = \boldsymbol{\mu}^\lambda$ yields $F(\boldsymbol{w}^\lambda, \boldsymbol{\mu}^\lambda) \geq F^*$. On the other hand, taking $\boldsymbol{w}' = \boldsymbol{w}^\lambda$ in the first display yields $F(\boldsymbol{w}^\lambda, \boldsymbol{\mu}^\lambda) \leq F^*$. Therefore $F(\boldsymbol{w}^\lambda, \boldsymbol{\mu}^\lambda) = F^*$, and the two displays imply

$$F(\boldsymbol{w}', \boldsymbol{\mu}^\lambda) \leq F(\boldsymbol{w}^\lambda, \boldsymbol{\mu}^\lambda) \leq F(\boldsymbol{w}^\lambda, \boldsymbol{\mu}') \qquad \forall \boldsymbol{w}' \in \Delta_K,\ \boldsymbol{\mu}' \in \Delta_{|\mathcal{X}|}.$$

Hence $(\boldsymbol{w}^\lambda, \boldsymbol{\mu}^\lambda) \in \mathcal{E}$, proving convexity.

For each $i \in \mathcal{K}_1$, by definition of $\mathcal{K}_1$ there exists a saddle point $(\boldsymbol{w}^{(i)}, \boldsymbol{\mu}^{(i)}) \in \mathcal{E}$ such that $w_i^{(i)} > 0$. Define the average pair

$$(\boldsymbol{w}^*, \boldsymbol{\mu}^*) \triangleq \frac{1}{|\mathcal{K}_1|} \sum_{i \in \mathcal{K}_1} (\boldsymbol{w}^{(i)}, \boldsymbol{\mu}^{(i)}).$$

By convexity of $\mathcal{E}$, we have $(\boldsymbol{w}^*, \boldsymbol{\mu}^*) \in \mathcal{E}$. If $j \in \mathcal{K}_0$, then $w_j = 0$ for every $(\boldsymbol{w}, \boldsymbol{\mu}) \in \mathcal{E}$, hence $w_j^* = 0$. If $j \in \mathcal{K}_1$, then

$$w_j^* = \frac{1}{|\mathcal{K}_1|} \sum_{i \in \mathcal{K}_1} w_j^{(i)} \ \geq \ \frac{1}{|\mathcal{K}_1|} w_j^{(j)} \ > \ 0,$$

so $w_j^* > 0$. □

Lemma 13 (Summability of harmonic compensators)**.** *Let $q_{n,i} = \mathbb{P}(I_{n+1} = i \mid \mathcal{H}_n)$. Then*

$$\sum_{n=0}^{\infty} \sum_{i=1}^{K} \frac{q_{n,i}}{N_{n,i}^2} \ < \ \infty \qquad \mathbb{P}_{\boldsymbol{\theta}}\text{-a.s.}$$

Proof. Fix $i \in [K]$ and define

$$Y_{n+1}^{(i)} \triangleq \frac{\mathbb{1}\{I_{n+1} = i\}}{N_{n,i}^2}, \qquad n \geq 0.$$

Since $N_{n,i}$ increases by 1 exactly on rounds with $I_{n+1} = i$, we have the deterministic bound

$$\sum_{n=0}^{\infty} Y_{n+1}^{(i)} = \sum_{n : I_{n+1} = i} \frac{1}{N_{n,i}^2} \leq \sum_{k=1}^{\infty} \frac{1}{k^2} \ < \ \infty \quad \text{a.s.}$$

Moreover,

$$\mathbb{E}\left[Y_{n+1}^{(i)} \mid \mathcal{H}_n\right] = \frac{q_{n,i}}{N_{n,i}^2}.$$

Define the martingale

$$M_m^{(i)} \triangleq \sum_{n=0}^{m-1} \left(Y_{n+1}^{(i)} - \mathbb{E}[Y_{n+1}^{(i)} \mid \mathcal{H}_n]\right), \qquad m \geq 1.$$

Its increments satisfy $|M_{n+1}^{(i)} - M_n^{(i)}| \leq 1/N_{n,i}^2$ and hence $\sum_{n=0}^{\infty} \mathbb{E}[(M_{n+1}^{(i)} - M_n^{(i)})^2 \mid \mathcal{H}_n] \leq \sum_{n=0}^{\infty} 1/N_{n,i}^4 < \infty$ a.s. Therefore $M_m^{(i)}$ converges almost surely to a finite limit.

Since

$$\sum_{n=0}^{m-1} \frac{q_{n,i}}{N_{n,i}^2} = \sum_{n=0}^{m-1} \mathbb{E}[Y_{n+1}^{(i)} \mid \mathcal{H}_n] = \sum_{n=0}^{m-1} Y_{n+1}^{(i)} - M_m^{(i)},$$

and the right-hand side converges almost surely to a finite limit, the series $\sum_n q_{n,i}/N_{n,i}^2$ converges almost surely. Summing over $i \in [K]$ completes the proof. □

Theorem 3. *Under Assumptions 1–7, let $\{\boldsymbol{w}_n\}_{n\geq 0}$ be the cost allocation sequence generated by Algorithm 1 and iteration* (11)*. Then,* $\min_{x\in\mathcal{X}} D_x^w(\boldsymbol{w}_n;\boldsymbol{\theta}_n)\to\Gamma^*$ *almost surely. As a consequence, for $\{\boldsymbol{p}_n\}_{n\geq 0}$ from Algorithm 1, we also have* $\min_{x\in\mathcal{X}} D_x(\boldsymbol{p}_n;\boldsymbol{\theta}_n)\to\Gamma^*$ *almost surely.*

Proof. Fix a saddle-point representative $(\boldsymbol{w}^*,\boldsymbol{\mu}^*)$ as in Lemma 12, and write $\mathcal{K}_1=\mathrm{Supp}(\boldsymbol{w}^*)$. Define the KL potential on the active face

$$V_n \triangleq D_{\mathrm{KL}}(\boldsymbol{w}^*\,\|\,\boldsymbol{w}_n)=\sum_{i\in\mathcal{K}_1} w_i^*\log\frac{w_i^*}{w_{n,i}}\;\geq\;0.$$

Since the forced-exploration schedule pulls every arm at least once, there exists a deterministic n_{init} such that $w_{n,i}>0$ for all $i\in\mathcal{K}_1$ and all $n\geq n_{\mathrm{init}}$, so $V_n<\infty$ for all $n\geq n_{\mathrm{init}}$.

Step 1: One-step drift bound on non-forced rounds. Fix $n\geq n_{\mathrm{init}}$ and define

$$u_{n+1,i}\triangleq\frac{w_{n+1,i}-w_{n,i}}{w_{n,i}}=\alpha_{n+1}\Big(\frac{\boldsymbol{e}_{I_{n+1}}(i)}{w_{n,i}}-1\Big),\qquad i\in\mathcal{K}_1,$$

so that $w_{n+1,i}=w_{n,i}(1+u_{n+1,i})$. Then

$$V_{n+1}-V_n=-\sum_{i\in\mathcal{K}_1} w_i^*\log(1+u_{n+1,i}).$$

For all sufficiently large n, we have $\alpha_{n+1}\leq c_{\max}/(c_{\min}(n+1))\leq 1/2$, hence $u_{n+1,i}\geq-\alpha_{n+1}\geq-1/2$ for every i. Using the inequality $-\log(1+u)\leq -u+u^2$ (valid for $u\geq -1/2$), we obtain

$$V_{n+1}-V_n\;\leq\;-\sum_{i\in\mathcal{K}_1} w_i^*u_{n+1,i}+\sum_{i\in\mathcal{K}_1} w_i^*u_{n+1,i}^2. \tag{48}$$

Taking conditional expectation given $\mathcal{H}_n$ yields

$$\mathbb{E}[V_{n+1}-V_n\mid\mathcal{H}_n]\;\leq\;-\widetilde{\alpha}_{n+1}\sum_{i\in\mathcal{K}_1} w_i^*\Big(\frac{\widetilde{h}_{n,i}}{w_{n,i}}-1\Big)+R_{n+1}, \tag{49}$$

where $\widetilde{\alpha}_{n+1}=\mathbb{E}[\alpha_{n+1}\mid\mathcal{H}_n]$, $\mathbb{E}[\alpha_{n+1}\boldsymbol{e}_{I_{n+1}}\mid\mathcal{H}_n]=\widetilde{\alpha}_{n+1}\widetilde{\boldsymbol{h}}_n$, and the quadratic remainder is

$$R_{n+1}\triangleq\mathbb{E}\left[\sum_{i\in\mathcal{K}_1} w_i^*u_{n+1,i}^2\;\Big|\;\mathcal{H}_n\right]\;\geq\;0.$$

On non-forced rounds,

$$\sum_{i\in\mathcal{K}_1} w_i^*\Big(\frac{\widetilde{h}_{n,i}}{w_{n,i}}-1\Big)=\sum_{i\in\mathcal{K}_1} w_i^*\Big(\frac{h_i^{x_n}(\boldsymbol{w}_n;\boldsymbol{\theta}_n)}{w_{n,i}}-1\Big)+\varepsilon_n, \tag{50}$$

where

$$\varepsilon_n\triangleq\sum_{i\in\mathcal{K}_1} w_i^*\left(\frac{\widetilde{h}_{n,i}}{w_{n,i}}-\frac{h_i^{x_n}(\boldsymbol{w}_n;\boldsymbol{\theta}_n)}{w_{n,i}}\right).$$

Using the IDS gradient representation $h_i^{x_n}(\boldsymbol{w}_n;\boldsymbol{\theta}_n)/w_{n,i}=[\nabla_{\boldsymbol{w}}D_{x_n}^w(\boldsymbol{w}_n;\boldsymbol{\theta}_n)]_i/D_{x_n}^w(\boldsymbol{w}_n;\boldsymbol{\theta}_n)$, define

$$\boldsymbol{g}_n\triangleq\nabla_{\boldsymbol{w}}D_{x_n}^w(\boldsymbol{w}_n;\boldsymbol{\theta}_n),\qquad D_n\triangleq D_{x_n}^w(\boldsymbol{w}_n;\boldsymbol{\theta}_n)>0.$$

Then the main term in (50) equals

$$S_n \triangleq \sum_{i\in\mathcal{K}_1} w_i^*\Big(\frac{[\boldsymbol{g}_n]_i}{D_n} - 1\Big) = \frac{\langle \boldsymbol{w}^* - \boldsymbol{w}_n, \boldsymbol{g}_n\rangle}{D_n}. \tag{51}$$

By concavity of $D_{x_n}^w(\cdot;\boldsymbol{\theta}_n)$,

$$D_{x_n}^w(\boldsymbol{w}^*;\boldsymbol{\theta}_n) \le D_{x_n}^w(\boldsymbol{w}_n;\boldsymbol{\theta}_n) + \langle \boldsymbol{g}_n, \boldsymbol{w}^* - \boldsymbol{w}_n\rangle = D_n + \langle \boldsymbol{g}_n, \boldsymbol{w}^* - \boldsymbol{w}_n\rangle,$$

so

$$S_n = \frac{\langle \boldsymbol{w}^* - \boldsymbol{w}_n, \boldsymbol{g}_n\rangle}{D_n} \;\ge\; \frac{D_{x_n}^w(\boldsymbol{w}^*;\boldsymbol{\theta}_n) - D_{x_n}^w(\boldsymbol{w}_n;\boldsymbol{\theta}_n)}{D_n} = \frac{D_{x_n}^w(\boldsymbol{w}^*;\boldsymbol{\theta}_n) - D_n}{D_n}. \tag{52}$$

Combining (49)–(52), we obtain on non-forced rounds

$$\mathbb{E}[V_{n+1} - V_n \mid \mathcal{H}_n] \;\le\; -\widetilde{\alpha}_{n+1}\frac{D_{x_n}^w(\boldsymbol{w}^*;\boldsymbol{\theta}_n) - D_n}{D_n} + \widetilde{\alpha}_{n+1}|\varepsilon_n| + R_{n+1}. \tag{53}$$

Step 2: Summability of error terms. By Proposition 7, there exists an a.s. finite n_0 such that $D_n = D_{x_n}^w(\boldsymbol{w}_n;\boldsymbol{\theta}_n) \ge D_*/2$ for all $n \ge n_0$. Together with Assumption 7, this implies $[\boldsymbol{g}_n]_i/D_n$ is uniformly bounded on $n \ge n_0$. By Lemma 2,

$$\left|\frac{\widetilde{h}_{n,i}}{w_{n,i}} - \frac{h_i^{x_n}(\boldsymbol{w}_n;\boldsymbol{\theta}_n)}{w_{n,i}}\right| \le \frac{c_{\max}}{nc_{\min}}\frac{h_i^{x_n}(\boldsymbol{w}_n,\boldsymbol{\theta}_n)}{w_{n,i}} = \frac{c_{\max}}{nc_{\min}}\frac{[\boldsymbol{g}_n]_i}{D_n} = O(1/n)$$

Hence $|\varepsilon_n| = O(1/n)$ implies $\sum_n \widetilde{\alpha}_{n+1}|\varepsilon_n| < \infty$ almost surely.

For the quadratic remainder, note that for $i \ne I_{n+1}$ we have $u_{n+1,i} = -\alpha_{n+1}$, while for $i = I_{n+1}$,

$$u_{n+1,I_{n+1}} = \alpha_{n+1}\Big(\frac{1}{w_{n,I_{n+1}}} - 1\Big) = \frac{C_{I_{n+1}}(\boldsymbol{\theta})}{B_{n+1}}\Big(\frac{B_n}{N_{n,I_{n+1}}C_{I_{n+1}}(\boldsymbol{\theta})} - 1\Big) \le \frac{1}{N_{n,I_{n+1}}}.$$

Therefore, for all n,

$$\sum_{i\in\mathcal{K}_1} w_i^* u_{n+1,i}^2 \;\le\; \alpha_{n+1}^2 + \frac{\mathbb{1}\{I_{n+1}\in\mathcal{K}_1\}}{N_{n,I_{n+1}}^2}.$$

Taking conditional expectations gives

$$R_{n+1} \le \mathbb{E}[\alpha_{n+1}^2 \mid \mathcal{H}_n] + \sum_{i\in\mathcal{K}_1}\frac{q_{n,i}}{N_{n,i}^2}.$$

Since $\sum_n \mathbb{E}[\alpha_{n+1}^2 \mid \mathcal{H}_n] \le \sum_n (c_{\max}/(c_{\min}(n+1)))^2 < \infty$ and Lemma 13 yields $\sum_n\sum_i q_{n,i}/N_{n,i}^2 < \infty$ almost surely, we conclude

$$\sum_{n=0}^{\infty} R_{n+1} \;<\; \infty \qquad \mathbb{P}_{\boldsymbol{\theta}}\text{-a.s.} \tag{54}$$

Step 3: Forced rounds contribute a summable perturbation. On forced rounds, I_{n+1} is $\mathcal{H}_n$-measurable, and for every $i \in \mathcal{K}_1$ with $i \ne I_{n+1}$ we have $w_{n+1,i} = (1-\alpha_{n+1})w_{n,i}$, hence

$$V_{n+1} - V_n = -\sum_{i\in\mathcal{K}_1} w_i^* \log\frac{w_{n+1,i}}{w_{n,i}} \le -\sum_{i\in\mathcal{K}_1\setminus\{I_{n+1}\}} w_i^*\log(1-\alpha_{n+1}) \le -\log(1-\alpha_{n+1}) \le 2\alpha_{n+1}$$

for all large n (so that $\alpha_{n+1} \le 1/2$). Since $\sum_{n:\mathsf{F}_n=1}\alpha_{n+1} < \infty$ under the forced schedule, the total forced-round contribution is almost surely summable.

Step 4: Robbins–Siegmund and identification of the limiting value. Let

$$v_n \triangleq \min_{x \in \mathcal{X}(\boldsymbol{\theta}_n)} D_x^w(\boldsymbol{w}_n; \boldsymbol{\theta}_n).$$

By Proposition 6, v_n converges almost surely to a finite limit v_∞. We show that $v_\infty = F^*$.

Suppose for contradiction that $v_\infty \le F^* - \delta$ on an event of positive probability, for some $\delta > 0$. Then for all sufficiently large n, we have $v_n \le F^* - \delta/2$ and, by detection consistency, $D_n = D_{x_n}^w(\boldsymbol{w}_n; \boldsymbol{\theta}_n) \le v_n + o(1) \le F^* - \delta/3$. On the other hand, since $\min_x D_x^w(\boldsymbol{w}^*; \boldsymbol{\theta}) = F^*$ and $\boldsymbol{\theta}_n \to \boldsymbol{\theta}$ almost surely, continuity implies $D_{x_n}^w(\boldsymbol{w}^*; \boldsymbol{\theta}_n) \ge F^* - \delta/6$ for all large n. Therefore, for all large n on this event,

$$D_{x_n}^w(\boldsymbol{w}^*; \boldsymbol{\theta}_n) - D_n \;\ge\; \delta/6.$$

Using (53) together with the summability of the error terms established above, we can write for all sufficiently large n,

$$\mathbb{E}[V_{n+1} \mid \mathcal{H}_n] \;\le\; V_n - U_n + W_n,$$

where

$$U_n \triangleq \mathsf{NF}_n \widetilde{\alpha}_{n+1} \frac{D_{x_n}^w(\boldsymbol{w}^*; \boldsymbol{\theta}_n) - D_n}{D_n} \ge 0,\; W_n \triangleq \widetilde{\alpha}_{n+1}|\varepsilon_n| + R_{n+1} \quad \text{and} \quad \sum_{n=0}^{\infty} W_n < \infty \text{ a.s.}$$

Robbins–Siegmund's almost-supermartingale theorem then implies $\sum_n U_n < \infty$ almost surely.

However, on the event $\{v_\infty \le F^* - \delta\}$ and for all large n, $D_n \le F^* + 1$ and $D_{x_n}^w(\boldsymbol{w}^*; \boldsymbol{\theta}_n) - D_n \ge \delta/6$, hence

$$U_n \;\ge\; \mathsf{NF}_n \widetilde{\alpha}_{n+1} \frac{\delta/6}{F^* + 1}.$$

Since $\widetilde{\alpha}_{n+1} \ge c_{\min}/(c_{\max}(n+1))$ and $\sum_n \mathsf{NF}_n/(n+1) = \infty$, we obtain $\sum_n U_n = \infty$ on this event, a contradiction. Therefore $\mathbb{P}(v_\infty \le F^* - \delta) = 0$ for every $\delta > 0$, i.e., $v_\infty \ge F^*$ almost surely.

Finally, by definition of $F^* = \max_{\boldsymbol{w} \in \Delta_K} \min_{x \in \mathcal{X}} D_x^w(\boldsymbol{w}; \boldsymbol{\theta})$ and since $\boldsymbol{\theta}_n \to \boldsymbol{\theta}$, we also have $v_n \le F^* + o(1)$, hence $v_\infty \le F^*$. Combining yields $v_\infty = F^*$ almost surely. This proves $\min_{x \in \mathcal{X}} D_x^w(\boldsymbol{w}_n; \boldsymbol{\theta}_n) \to F^*$ almost surely.

Step 5: Transfer from cost allocations to sample allocations. For each $x \in \mathcal{X}(\boldsymbol{\theta}_n)$,

$$D_x(\boldsymbol{p}_n; \boldsymbol{\theta}_n) = \frac{\bar{C}_{\boldsymbol{\theta}}(\boldsymbol{p}_n)}{\bar{C}_{\boldsymbol{\theta}_n}(\boldsymbol{p}_n)} \inf_{\boldsymbol{\vartheta} \in \mathrm{Alt}_x(\boldsymbol{\theta}_n)} \sum_{i=1}^{K} w_{n,i} \frac{\mathrm{KL}(P_{\boldsymbol{\theta}_n,i} \| P_{\boldsymbol{\vartheta},i})}{C_i(\boldsymbol{\theta})} = \frac{\bar{C}_{\boldsymbol{\theta}}(\boldsymbol{p}_n)}{\bar{C}_{\boldsymbol{\theta}_n}(\boldsymbol{p}_n)} D_x^w(\boldsymbol{w}_n; \boldsymbol{\theta}_n).$$

By continuity of $C_i(\cdot)$ and $\boldsymbol{\theta}_n \to \boldsymbol{\theta}$, the prefactor $\bar{C}_{\boldsymbol{\theta}}(\boldsymbol{p}_n)/\bar{C}_{\boldsymbol{\theta}_n}(\boldsymbol{p}_n) \to 1$. Therefore $\min_x D_x(\boldsymbol{p}_n; \boldsymbol{\theta}_n) - \min_x D_x^w(\boldsymbol{w}_n; \boldsymbol{\theta}_n) \to 0$ almost surely, and hence $\min_x D_x(\boldsymbol{p}_n; \boldsymbol{\theta}_n) \to F^*$ almost surely. □

Remark 4 (Boundary separation on potentially active arms)**.** *When the saddle-point set is not a singleton, different algorithms may converge to different saddle points. For instance, it may happen that $\mathcal{E}_{p_i} = [0, 1/2]$ for some coordinate i, so both boundary ($p_i = 0$) and interior ($p_i > 0$) saddle points exist. The KL-potential argument used in the proof of Theorem 3 implies a simple but useful* boundary-separation *property for coordinates that are positive in a saddle-point representative.*

Let $(\boldsymbol{w}^*, \boldsymbol{\mu}^*)$ be chosen as in Lemma 12, so that $(\boldsymbol{w}^*)_i > 0$ for all $i \in \mathcal{K}_1$, and recall the KL potential on the active face

$$V_n \triangleq D_{\mathrm{KL}}(\boldsymbol{w}^* \| \boldsymbol{w}_n) = \sum_{i \in \mathcal{K}_1} w_i^* \log \frac{w_i^*}{w_{n,i}}.$$

Robbins–Siegmund yields that V_n converges almost surely in Theorem 3; in particular, V_n is almost surely bounded, say $\sup_n V_n \le \bar{V} < \infty$ on an event of probability one. Fix any $i \in \mathcal{K}_1$. Since each term in the sum defining V_n is nonnegative, we have

$$w_i^* \log \frac{w_i^*}{w_{n,i}} \;\le\; V_n \;\le\; \bar{V} \qquad \forall n,$$

hence $\log \frac{w_i^*}{w_{n,i}} \le \bar{V}/w_i^*$ and therefore

$$w_{n,i} \;\ge\; w_i^* \exp\Big(-\frac{\bar{V}}{w_i^*}\Big) \;>\; 0 \qquad \forall n.$$

Consequently, $\inf_n w_{n,i} > 0$ for all $i \in \mathcal{K}_1$, and every accumulation point of $(\boldsymbol{w}_n)$ assigns strictly positive mass to all coordinates in $\mathcal{K}_1$. Equivalently, each potentially active arm (i.e., one that is positive in some saddle point) receives a linear *number of samples under the* `PAN`*/IDS dynamics started from the interior.*

We view this as a robustness feature of IDS-style sampling dynamics: it prevents the algorithm from selecting sparse boundary saddle points that would "starve" potentially active arms (e.g., $p_i \to 0$, leading to sublinear sampling).

F Examples

We illustrate two simple pathologies in linear best-arm identification (BAI) under unit noise variance and unit costs. Throughout, the mean of arm i is $m_i \triangleq \langle \boldsymbol{a}_i, \boldsymbol{\theta} \rangle$, and the optimal arm is $I^* \triangleq \arg\max_{i \in [K]} m_i$ (assumed unique). For $j \neq I^*$, define the gap and contrast direction

$$\Delta_j \triangleq m_{I^*} - m_j, \qquad \boldsymbol{u}_j \triangleq \boldsymbol{a}_{I^*} - \boldsymbol{a}_j.$$

Given an allocation $\boldsymbol{p} \in \Delta_K$, let

$$V_{\boldsymbol{p}} \triangleq \sum_{i=1}^{K} p_i \boldsymbol{a}_i \boldsymbol{a}_i^\top$$

denote the (normalized) design matrix. In Gaussian linear BAI with unit variances, the usual Chernoff-type information rate against rival j takes the form

$$D_j(\boldsymbol{p}; \boldsymbol{\theta}) \triangleq \frac{\Delta_j^2}{2 \boldsymbol{u}_j^\top V_{\boldsymbol{p}}^{-1} \boldsymbol{u}_j}, \qquad j \neq I^*, \tag{55}$$

and the maximin value is $\Gamma^* \triangleq \max_{\boldsymbol{p} \in \Delta_K} \min_{j \neq I^*} D_j(\boldsymbol{p}; \boldsymbol{\theta})$.

F.1 Uniform allocation can be arbitrarily suboptimal

Consider a $d = 2$ linear bandit with parameter $\boldsymbol{\theta} = (1, -C)^\top$ for some $C > 0$. Let $K \geq 2$ and define arms

$$\boldsymbol{a}_1 = \boldsymbol{e}_1, \qquad \boldsymbol{a}_2 = \boldsymbol{e}_1 + \varepsilon \boldsymbol{e}_2, \qquad \boldsymbol{a}_k = \alpha \boldsymbol{e}_1, \quad k = 3, \dots, K,$$

where $\alpha \in (0, 1)$ is fixed and $0 < \varepsilon < (1 - \alpha)/C$. The means satisfy

$$m_1 = 1, \qquad m_2 = 1 - C\varepsilon, \qquad m_k = \alpha, \quad k \geq 3,$$

so $I^* = 1$ and arm 2 is the unique near competitor:

$$\Delta_2 = C\varepsilon, \qquad \Delta_k = 1 - \alpha > \Delta_2, \quad k \geq 3.$$

The associated contrast directions are

$$\boldsymbol{u}_2 = \boldsymbol{a}_1 - \boldsymbol{a}_2 = -\varepsilon \boldsymbol{e}_2, \qquad \boldsymbol{u}_k = \boldsymbol{a}_1 - \boldsymbol{a}_k = (1 - \alpha)\boldsymbol{e}_1, \quad k \geq 3.$$

Fix $\boldsymbol{p} \in \Delta_K$ and write $S \triangleq \sum_{k=3}^K p_k$ and $s \triangleq p_1 + p_2 + \alpha^2 S$. A direct calculation gives

$$V_{\boldsymbol{p}} = \begin{bmatrix} s & \varepsilon p_2 \\ \varepsilon p_2 & \varepsilon^2 p_2 \end{bmatrix}, \qquad \det(V_{\boldsymbol{p}}) = \varepsilon^2 p_2 (s - p_2) = \varepsilon^2 p_2 (p_1 + \alpha^2 S),$$

and hence

$$V_{\boldsymbol{p}}^{-1} = \frac{1}{\det(V_{\boldsymbol{p}})} \begin{bmatrix} \varepsilon^2 p_2 & -\varepsilon p_2 \\ -\varepsilon p_2 & s \end{bmatrix}.$$

Plugging into (55) yields

$$\begin{aligned} D_2(\boldsymbol{p}; \boldsymbol{\theta}) &= \frac{(C\varepsilon)^2}{2} \cdot \frac{p_2(s - p_2)}{s} = \frac{(C\varepsilon)^2}{2} \cdot \frac{p_2(p_1 + \alpha^2 S)}{p_1 + p_2 + \alpha^2 S}, \\ D_k(\boldsymbol{p}; \boldsymbol{\theta}) &= \frac{(1 - \alpha)^2}{2} \cdot \frac{1}{(1 - \alpha)^2 \boldsymbol{e}_1^\top V_{\boldsymbol{p}}^{-1} \boldsymbol{e}_1} = \frac{s - p_2}{2} = \frac{p_1 + \alpha^2 S}{2}, \qquad k \geq 3. \end{aligned}$$

Now consider the uniform allocation $p_i = 1/K$. Then $S = (K - 2)/K$ and

$$s = \frac{2 + \alpha^2 (K - 2)}{K}, \qquad s - p_2 = \frac{1 + \alpha^2 (K - 2)}{K}.$$

Therefore

$$D_2^{\text{unif}} = \frac{(C\varepsilon)^2}{2} \cdot \frac{\frac{1}{K} \cdot \frac{1 + \alpha^2 (K-2)}{K}}{\frac{2 + \alpha^2 (K-2)}{K}} = \frac{(C\varepsilon)^2}{2} \cdot \frac{1 + \alpha^2 (K - 2)}{K[2 + \alpha^2 (K - 2)]} = \Theta\left(\frac{(C\varepsilon)^2}{K} \right),$$

whereas for $k \geq 3$,

$$D_k^{\text{unif}} = \frac{s - p_2}{2} = \frac{1 + \alpha^2 (K - 2)}{2K} \longrightarrow \frac{\alpha^2}{2}.$$

Hence the bottleneck under uniform sampling is the hard rival 2:

$$\min_{j \neq 1} D_j^{\text{unif}} = D_2^{\text{unif}} = \Theta\left(\frac{(C\varepsilon)^2}{K} \right).$$

By contrast, allocating only between arms 1 and 2 (set $p_1 = p_2 = \frac{1}{2}$ and $p_k = 0$ for $k \geq 3$) gives

$$D_2(\boldsymbol{p}; \boldsymbol{\theta}) = \frac{(C\varepsilon)^2}{8}, \qquad D_k(\boldsymbol{p}; \boldsymbol{\theta}) = \frac{1}{4}, \quad k \geq 3,$$

so for ε sufficiently small we have $\min_{j \neq 1} D_j(\boldsymbol{p}; \boldsymbol{\theta}) = \frac{(C\varepsilon)^2}{8}$ and thus $\Gamma^* \geq (C\varepsilon)^2/8$. Consequently,

$$\frac{\Gamma^*}{\min_{j \neq 1} D_j^{\text{unif}}} \gtrsim \frac{(C\varepsilon)^2/8}{\Theta((C\varepsilon)^2/K)} = \Omega(K) \xrightarrow[K \to \infty]{} \infty.$$

Interpretation. All distractors $k \geq 3$ lie on the $\boldsymbol{e}_1$ axis and contribute no direct information in the $\boldsymbol{e}_2$ direction that separates arms 1 and 2 (since $\boldsymbol{u}_2 = -\varepsilon \boldsymbol{e}_2$). Uniform sampling wastes a $1 - 2/K$ fraction of samples on these distractors, forcing the worst-case information rate to decay as $1/K$, whereas an optimal design concentrates on arms 1 and 2 and maintains a $\Theta((C\varepsilon)^2)$ rate.

F.2 β-tuned top-two sampling can be arbitrarily suboptimal

We next exhibit a simple instance where a fixed-β top-two rule can be arbitrarily worse than the maximin (IDS-style) allocation. Consider $d = 2$, $K = 4$, parameter $\boldsymbol{\theta} = (-1, -2)^\top$, and arms

$$\boldsymbol{a}_1 = \boldsymbol{e}_1, \qquad \boldsymbol{a}_2 = \boldsymbol{e}_2, \qquad \boldsymbol{a}_3 = l\boldsymbol{e}_1, \qquad \boldsymbol{a}_4 = \boldsymbol{e}_1 + l\boldsymbol{e}_2,$$

where $l > 1$. Then $I^* = 1$ with gaps

$$\Delta_2 = 1, \qquad \Delta_3 = l - 1, \qquad \Delta_4 = 2l,$$

and contrast directions

$$\boldsymbol{u}_2 = (1, -1)^\top, \qquad \boldsymbol{u}_3 = (1 - l, 0)^\top, \qquad \boldsymbol{u}_4 = (0, -l)^\top.$$

For $\boldsymbol{p} \in \Delta_4$, the design matrix is

$$V_{\boldsymbol{p}} = \begin{bmatrix} p_1 + l^2 p_3 + p_4 & l p_4 \\ l p_4 & p_2 + l^2 p_4 \end{bmatrix}, \qquad \det(V_{\boldsymbol{p}}) = l^4 p_3 p_4 + l^2(p_1 p_4 + p_2 p_3) + p_1 p_2 + p_2 p_4,$$

and

$$V_{\boldsymbol{p}}^{-1} = \frac{1}{\det(V_{\boldsymbol{p}})} \begin{bmatrix} p_2 + l^2 p_4 & -l p_4 \\ -l p_4 & p_1 + l^2 p_3 + p_4 \end{bmatrix}.$$

A direct substitution into (55) yields

$$D_2(\boldsymbol{p}; \boldsymbol{\theta}) = \frac{\det(V_{\boldsymbol{p}})}{2(p_1 + p_2 + l^2 p_3 + (l+1)^2 p_4)},$$
$$D_3(\boldsymbol{p}; \boldsymbol{\theta}) = \frac{\det(V_{\boldsymbol{p}})}{2(p_2 + l^2 p_4)}, \qquad D_4(\boldsymbol{p}; \boldsymbol{\theta}) = \frac{2\det(V_{\boldsymbol{p}})}{p_1 + l^2 p_3 + p_4}.$$

One can verify that $D_2(\boldsymbol{p}; \boldsymbol{\theta}) \leq \min\{D_3(\boldsymbol{p}; \boldsymbol{\theta}), D_4(\boldsymbol{p}; \boldsymbol{\theta})\}$ for all $\boldsymbol{p} \in \Delta_4$, so the maximin value is $\Gamma^* = \max_{\boldsymbol{p} \in \Delta_4} D_2(\boldsymbol{p}; \boldsymbol{\theta})$. In particular, the feasible allocation $\boldsymbol{p}^\dagger = (0, 0, \frac{1}{2}, \frac{1}{2})$ gives the lower bound

$$\Gamma^* \geq D_2(\boldsymbol{p}^\dagger; \boldsymbol{\theta}) = \frac{l^4}{4(2l^2 + 2l + 1)} = \Theta(l^2).$$

By contrast, a top-two Thompson sampling rule with fixed tuning $\beta = \frac{1}{2}$ concentrates asymptotically on the leader and its closest challenger in mean, which in this instance yields the limiting allocation

$$\boldsymbol{p}^{\mathrm{TT}} = \left(\tfrac{1}{2}, \tfrac{1}{2}, 0, 0\right).$$

For this allocation, $V_{\boldsymbol{p}^{\mathrm{TT}}} = \frac{1}{2} I_2$, so

$$\min_{j \neq 1} D_j(\boldsymbol{p}^{\mathrm{TT}}; \boldsymbol{\theta}) = D_2(\boldsymbol{p}^{\mathrm{TT}}; \boldsymbol{\theta}) = \frac{1}{8}.$$

Consequently,

$$\frac{\Gamma^*}{\min_{j \neq 1} D_j(\boldsymbol{p}^{\mathrm{TT}}; \boldsymbol{\theta})} \;\geq\; 8 D_2(\boldsymbol{p}^{\dagger}; \boldsymbol{\theta}) = \Theta(l^2) \;\xrightarrow[l \to \infty]{}\; \infty,$$

showing that a fixed-β top-two rule can be arbitrarily suboptimal relative to an IDS-style allocation that targets the maximin information rate.

G Additional Numerical Examples

We consider a simple $d = 3$, $K = 6$ linear instance in which the best arm is $a_1 = e_1$ and the closest competitor is $a_2 = e_1 + \frac{1}{2} e_2$ under $\boldsymbol{\theta} = (1, -0.02, 0)$, so the gap is only $\langle a_1 - a_2, \boldsymbol{\theta} \rangle = 0.01$ and identifying the best arm requires accurately estimating the *nuisance* coordinate θ_2. Crucially, there are two arms aligned with e_2, namely $a_3 = a_4 = e_2$, but with different noise levels: arm 3 has low variance $\sigma_3^2 = 0.04$ while arm 4 has variance 1.

In the *equal-cost* version (all $c_i \equiv 1$), the low-variance arm 3 is genuinely preferable, and LinGapE behaves similarly to PAN and Track-and-Stop: the median stopping times are 460 (PAN), 450 (Track-and-Stop), and 595 (LinGapE) pulls (Figure 2, Left Panel).

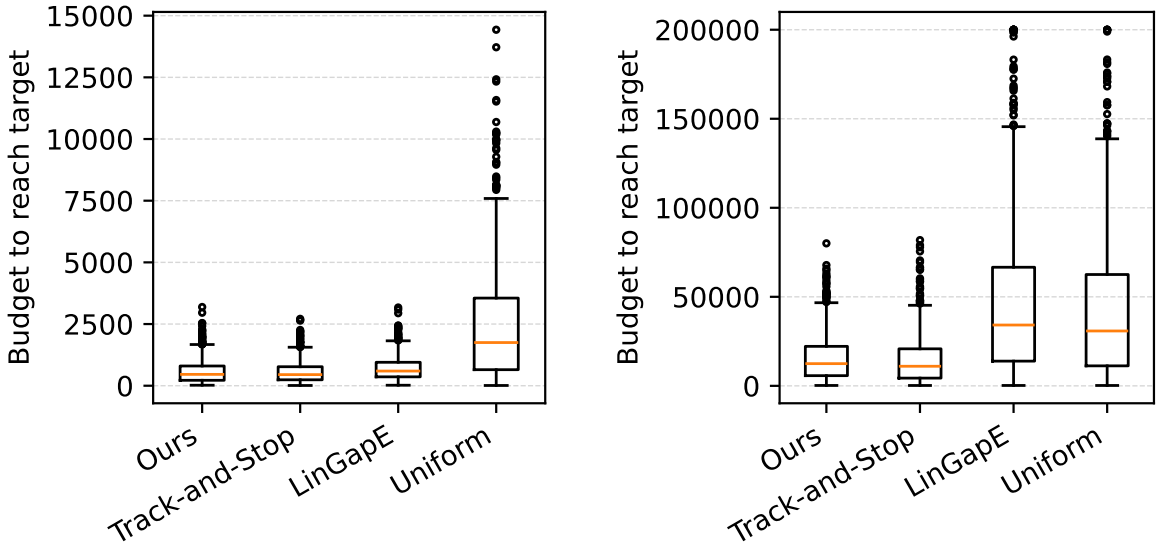

Figure 2: Budget needed to reach posterior threshold. Left: uniform cost. Right: Heterogeneous cost.

We then modify *only* one number: we set the cost of arm 3 to $c_3 = 100$ while leaving the arms, $\boldsymbol{\theta}$, and variances unchanged. This creates a "cost trap": arm 3 is 25× less noisy but 100× more expensive, so it is *four times worse* per unit budget than its cheap duplicate arm 4. A cost-aware method should therefore switch almost entirely to sampling arm 4 to learn θ_2 efficiently in budget.

The results are collected in Figure 2, Right Panel. Empirically, PAN and Track-and-Stop remain stable, reaching the target posterior error after about 1.25×10^4 and 1.10×10^4 budget units in median, respectively. In contrast, a naive (cost-agnostic) LinGapE implementation becomes dramatically cost-inefficient: its median budget to target jumps to 3.42×10^4 (about a 3× increase relative to PAN/Track-and-Stop and $\sim 60\times$ relative to its equal-cost median when "budget" coincides with pulls), with a heavy tail (95th percentile 1.28×10^5) and occasional failures to reach the target within the budget cap (0.7% of runs).

This experiment highlights that heterogeneous costs can qualitatively change the optimal allocation, and that plugging a standard LinGapE rule into a cost-budgeted setting without explicitly incorporating costs can lead to a severe performance collapse even when the underlying geometry and noise model are unchanged.

H Auxiliary facts

In this section, we compile several classical results on set-valued maps and differential inclusions that facilitate our proofs.

Envelope theorem. Let X be a choice set and let $t \in [0,1]$ be the relevant parameter. Consider the parameterized objective function $f : X \times [0,1] \to \mathbb{R}$, and define the value function $v : [0,1] \to \mathbb{R}$ and the optimal choice correspondence (set-valued function) X^* by

$$v(t) = \sup_{x \in X} f(x,t) \quad \text{and} \quad X^*(t) = \{x \in X : f(x,t) = v(t)\}.$$

The Envelope theorem, Milgrom and Segal (2002, Theorem 1), states that: Assume that for a given $t \in [0,1]$ and for some $x^* \in X^*(t)$ the partial derivative $f_t(x^*, t)$ exists. If v is differentiable at t, then $v'(t) = f_t(x^*, t)$.

Upper hemicontinuity for compact-valued correspondence. Beavis and Dobbs (1990, Theorem 3.2) provide an equivalent characterization of upper hemicontinuity for compact-valued set-valued map $F : X \rightrightarrows Y$. F is upper hemicontinuous at $x \in X$ if, and only if, for every sequence $\{x_n\}$ converging to x and every sequence $\{y_n\}$ with $y_n \in F(x_n)$, there exists a converging subsequence of $\{y_n\}$ whose limit belongs to $F(x)$.

Berge's Maximum Theorem. Below is a rephrase of Beavis and Dobbs (1990, Theorem 3.6). Let $X \subset \mathbb{R}^m$, $Y \subset \mathbb{R}^k$ and $\Xi : X \rightrightarrows Y$ be a set-valued map with nonempty, compact values. Let $f : X \times Y \to \mathbb{R}$ be a continuous function. Define the set-valued function $M : X \rightrightarrows Y$, the maximizers $M(x) \triangleq \arg\max_{y \in \Xi(x)} f(x,y)$, and the corresponding value function $v : X \to \mathbb{R}$ by

$v(x) = \max_{y \in \Xi(x)} f(x, y)$. If Ξ is continuous at x, then v is continuous at x and the set-valued function M is closed, compact-valued and upper hemicontinuous at x.

Existence of solutions to a differential inclusion. Consider the autonomous differential inclusion

$$\dot{\boldsymbol{x}}(t) \in F(\boldsymbol{x}(t)), \qquad \boldsymbol{x}(0) = \boldsymbol{x}_0,$$

where $F : \mathbb{R}^m \rightrightarrows \mathbb{R}^m$ satisfies: (i) $\text{graph}(F) = \{(\boldsymbol{x}, \boldsymbol{y}) : \boldsymbol{y} \in F(\boldsymbol{x})\}$ is closed; (ii) $F(\boldsymbol{x})$ is nonempty, compact, and convex for every $\boldsymbol{x} \in \mathbb{R}^m$; (iii) there exists $c > 0$ such that $\sup_{\boldsymbol{z} \in F(\boldsymbol{x})} \|\boldsymbol{z}\| \leq c(1 + \|\boldsymbol{x}\|)$ for all $\boldsymbol{x} \in \mathbb{R}^m$. Then, for every initial condition $\boldsymbol{x}_0 \in \mathbb{R}^m$, there exists an absolutely continuous trajectory $\boldsymbol{x} : [0, \infty) \to \mathbb{R}^m$ such that $\dot{\boldsymbol{x}}(t) \in F(\boldsymbol{x}(t))$ for almost every $t \geq 0$. See, e.g., Aubin and Cellina (1984, Chapter 2.1).

Existence of measurable selections. Aubin and Cellina (1984, Corollary 1, Section 1.14) states that, let $f : X \times U \to X$ be continuous, where U is a compact separable metric space. Assume that there exist an interval I and an absolutely continuous function $x : I \to \mathbb{R}^n$, such that

$$x'(t) \in f(x(t), U) \quad \text{for almost every } t \in I.$$

Then, there exists a Lebesgue measurable function $u : I \to U$ such that

$$x'(t) = f(x(t), u(t)) \quad \text{for almost every } t \in I.$$