BEST-ARM IDENTIFICATION IN FACTORED MULTI-AGENT BANDITS VIA JUNCTION TREES

Anonymous authors
Paper under double-blind review

ABSTRACT

Factored Multi-Agent Multi-Armed Bandits (MAMABs) model cooperative decision problems with exponentially large joint-action spaces and sparse local interactions. In fixed-confidence Best-Arm Identification (BAI), computing the information-optimal sampling allocation remains combinatorial. We derive a Junction-Tree (JT) representation of the characteristic-time program: locally consistent clique marginals characterize feasible sampling allocations, while a gap-augmented dynamic program handles competing joint actions without enumerating the combinatorial action space. The resulting complexity depends on the induced width of the graphical decomposition and a gap-state width measuring the number of cumulative reward gaps retained during elimination. For lattice-valued means, the resulting dynamic program is exact with pseudo-polynomial complexity; for arbitrary real-valued rewards, quantizing this gap dimension yields a certified approximation with a controllable accuracy-complexity trade-off. We further introduce JT Track-and-Stop, (JT-TaS) a δ-correct algorithm that asymptotically attains the instance-dependent lower bound. Experiments recover the exact optimum on enumerable instances and scale to instances with large joint action spaces.

1 INTRODUCTION

Factored Multi-Agent Multi-Armed Bandits (MAMABs) model cooperative decision problems in which N agents select actions jointly and the global reward decomposes according to a factor graph (Amin et al., 2011; Bargiacchi et al., 2018). Each agent chooses one of K actions and the reward is the sum of ρ local factors, with factor e depending only on the actions in its scope $\mathcal{S}_e \subseteq [N]$. A joint action belongs to a space of size K^N. When a joint action is played, the individual factor rewards are observed separately, yielding *semi-bandit feedback* (Kveton et al., 2015). This structure arises naturally in distributed control problems such as wind-farm or radio-network optimization (Bargiacchi et al., 2018; Vannella et al., 2023a), with globally coupled decisions and sparse interactions.

Best-Arm Identification (BAI) with fixed confidence aims to identify the action with maximal expected reward with probability at least $1-\delta$, while minimizing the number of samples (Mannor and Tsitsiklis, 2004). When the action set is large, exploiting the instance structure is essential for both statistical and computational efficiency (Jourdan et al., 2021). The asymptotic sample-complexity lower bound is governed by an instance-dependent *characteristic time* $T^\star_\theta$, defined through an information-allocation problem whose optimizer prescribes how observations should be allocated to distinguish the optimal action from its most confusing alternatives (Garivier and Kaufmann, 2016). Computing this optimal allocation can itself become the dominant computational bottleneck in structured action spaces.

For factored MAMABs, the lower bound and its optimal allocations are known (Vannella et al., 2023a; Jourdan et al., 2021), but computing it requires solving an hard combinatorial program. Mean-Field (MF) approximations can restores tractability by restricting the feasible allocation to product-form sampling distributions Vannella et al. (2023a), but typically can yield loose approximations. Generic combinatorial semi-bandit methods avoid this restriction and can attain asymptotic optimality through oracle-based optimization (Jourdan et al., 2021; Tzeng et al., 2023). However, their computational guarantees rely on generic combinatorial optimization oracles rather than characterizing when the information-allocation problem itself becomes tractable from the sparsity of the factor graph. This motivates a structure-dependent treatment of the characteristic time.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

We address this problem by exploiting the connection between factored rewards and exact inference in graphical models (Koller and Friedman, 2009; Wainwright and Jordan, 2008). For an elimination order O, we construct a junction tree (JT) whose bags contain at most $w_O + 1$ agents, where w_O is its *induced width*. The JT resolves the two exponential objects appearing in $T_\theta^\star$. The largest local configuration then involves at most $w_O + 1$ agents. On the same JT, locally consistent clique marginals characterize exactly the globally feasible sampling allocations, while optimization over competing joint actions is performed by augmenting the local computation with their cumulative reward gap. The latter introduces a second complexity quantity, the *gap-state width* $H_\theta(O)$, defined as the largest number of distinct cumulative gap values retained during elimination.

Contributions. Our main contributions are as follows:

1. **JT representation of information-optimal exploration.** For any elimination order O, we obtain a unified local representation of both feasible sampling allocations and competing joint actions. The resulting characteristic-time program is equivalent to the original one and has $O\left((NH_\theta(O) + \rho)K^{w_O+1}\right)$ variables and $O\left((NH_\theta(O)^2 + \rho)K^{w_O+1}\right)$ constraints.
2. **Certified accuracy-complexity trade-off.** For lattice-valued local means, the characteristic time can be computed exactly with pseudo-polynomial complexity. For arbitrary real-valued rewards, we quantize the local means to obtain a finite gap-state representation with a deterministic approximation guarantee: for any $\varepsilon \in (0, 1]$, the resulting computable upper estimate of the characteristic time is within a factor $1 + \varepsilon$ of $T_\theta^\star$. A grid resolution of order $\varepsilon\Delta_{\min}/\rho$ suffices, yielding an explicit accuracy complexity trade-off.
3. **Asymptotically optimal JT-TaS.** We introduce *Junction-Tree Track-and-Stop* (JT-TaS), which learns directly in the JT representation: quantized empirical instances define increasingly accurate information-optimal bag marginals, ancestral JT sampling realizes these marginals as globally consistent joint actions, and a conservative Generalized Log-likelihood Ratio (GLR) provides tractable stopping evidence. JT-TaS is δ-PAC and asymptotically attains the instance-specific lower bound. Experiments quantify the resulting computational-statistical trade-offs.

2 Related Work

Factored multi-agent bandits. Factored MAMABs have been studied for regret minimization (Stranders et al., 2012; Bargiacchi et al., 2018; Verstraeten et al., 2020; Jin et al., 2024; Vannella et al., 2023b), and for BAI (Bargiacchi et al., 2022; Vannella et al., 2023a). The closest work, Vannella et al. (2023a), derives the fixed-confidence BAI lower bound used in our work, and characterizes its optimal allocation, but resorts to a MF approximation of the characteristic time (an upper bound) to avoid the combinatorial exact optimization. We instead solve the lower bound optimization time by directly exploiting the true graphical structure of the model.

Combinatorial BAI. Our work builds on the instance-dependent BAI lower-bound literature, where optimal algorithms track instance-dependent sampling proportions (Garivier and Kaufmann, 2016; Wang et al., 2021). It is closely related to combinatorial BAI and semi-bandits, where exponentially large action sets are handled through optimization oracles (Chen et al., 2014; Jourdan et al., 2021; Tzeng et al., 2023). Factored MAMABs induce a particular semi-bandit structure in which local factor configurations must be mutually compatible on overlapping agent variables. Rather than treating this structure through a generic oracle, we exploit a JT representation explicitly. The arithmetic dependence of our gap-state dynamic program is also related to budgeted combinatorial optimization techniques used in Cuvelier et al. (2021).

Graphical bandits and inference. Graphical structure has previously been used to obtain tractable algorithms for large-action-space bandits (Amin et al., 2011; Yu et al., 2020). Our computational construction builds on Variable Elimination (VE) and JT methods (Lauritzen and Spiegelhalter, 1988a; Wainwright and Jordan, 2008), which replace global computation by local operations over bags and separators. JT consistency represents globally realizable factor allocations exactly, and augment the corresponding messages with the cumulative signed gap required by the BAI objective, in the spirit of constrained message passing with global additive statistics (Bauer et al., 2023). In our setting, these ideas yield an exact structure-dependent formulation of the characteristic-time optimization, with complexity jointly controlled by the induced width and the gap state width.

3 Problem Setting

We consider a factored MAMAB model with N agents, indexed by $\mathcal{S} = [N] \triangleq \{1, \dots, N\}$. Agent i can select actions from a finite set $\mathcal{A}_i$, and we assume $|\mathcal{A}_i| = K \geq 2$ for all $i \in [N]$. A global action is defined as $a = (a_1, \dots, a_N) \in \mathcal{A}$, where $\mathcal{A} \triangleq \times_{i \in [N]} \mathcal{A}_i$, so that $|\mathcal{A}| = A = K^N$.

Factored reward model. The reward function admits a factorized structure over ρ possibly overlapping groups of agents $(\mathcal{S}_e)_{e \in [\rho]}$, where $\mathcal{S}_e \subseteq \mathcal{S}$ and $|\mathcal{S}_e| = N_e$. For each factor $e \in [\rho]$, we denote by $\mathcal{A}_e = \times_{i \in \mathcal{S}_e} \mathcal{A}_i$ the corresponding factor action space, and by $a_e = (a_i)_{i \in \mathcal{S}_e} \in \mathcal{A}_e$ the restriction of a global action a to $\mathcal{S}_e$. Its cardinality is $A_e = |\mathcal{A}_e| = K^{N_e}$, and we define $\widetilde{A} = \sum_{e \in [\rho]} A_e$. Each factor e is associated with an unknown factor mean-reward function $\theta_e : \mathcal{A}_e \to \mathbb{R}$, and we denote the complete bandit instance by $\theta = (\theta_e(x))_{e \in [\rho], x \in \mathcal{A}_e} \in \mathbb{R}^{\widetilde{A}}$. When a global action a is played, factor e produces an observation $r_e(a_e) \sim \mathcal{N}(\theta_e(a_e), \sigma^2)$, where $\sigma^2 > 0$ is known. The global reward is $r(a) = \sum_{e \in [\rho]} r_e(a_e)$, with mean $\theta(a) = \mathbb{E}[r(a)] = \sum_{e \in [\rho]} \theta_e(a_e)$.

Sequential decision process. At each round $t \geq 1$, the learner selects a global action $a_t = (a_{t,1}, \dots, a_{t,N}) \in \mathcal{A}$ and observes the vector of factor rewards $r_t = (r_{t,1}, \dots, r_{t,\rho})$. Rewards are independent across factors and rounds, conditional on the selected action. The action a_{t+1} may depend on the history $\mathcal{H}_t = (a_s, r_s)_{s \in [t]}$. Since the individual factor rewards are observed separately, rather than only through their sum, this interaction corresponds to *semi-bandit feedback*.

Best-Arm Identification. We investigate fixed confidence BAI in MAMABs. Let $a^\star \in \arg\max_{a \in \mathcal{A}} \theta(a)$ be the best global action, which we assume to be unique. A BAI algorithm consists of (i) a *sampling rule* a_t, which adaptively selects global actions based on the history of observations $\mathcal{H}_{t-1}$, (ii) a *stopping rule* τ, determining when data collection terminates, and (iii) a *recommendation rule* $\hat{a}_\tau$, returning an estimate of the best global action. The objective is to identify $a^\star$ with probability at least $1 - \delta$ while minimizing the expected number of rounds $\mathbb{E}_\theta[\tau]$. Let Θ denote the set of bandit instances with a unique optimal global action. We focus on the class of δ-Probably Approximately Correct (PAC) algorithms, according to the following definition.

Definition 1. Let $\delta \in (0, 1/2)$. An algorithm is δ-PAC if $\forall \theta$, $\mathbb{P}_\theta(\hat{a}_\tau \neq a^\star) \leq \delta$ and $\mathbb{P}_\theta(\tau < \infty) = 1$.

4 Sample-Complexity Lower Bound

We recall the instance-dependent lower bound for fixed-confidence BAI in factored MAMABs. Its optimizer specifies the sampling allocation targeted by asymptotically optimal algorithms and provides the starting point for our development.

Notations. Let $\Lambda = \{w \in \mathbb{R}_+^{|\mathcal{A}|} : \sum_{a \in \mathcal{A}} w_a = 1\}$ denote the simplex of distributions over global actions. Any $w \in \Lambda$ induces factor-level marginals $\tilde{w}_{e,x} = \sum_{a \in \mathcal{A}: a_e = x} w_a$ for each factor $e \in [\rho]$ and local action $x \in \mathcal{A}_e$. The set of all realizable factor allocations is

$$\tilde{\Lambda} = \left\{ \tilde{w} : \exists w \in \Lambda : \tilde{w}_{e,x} = \sum_{a \in \mathcal{A}: a_e = x} w_a, \forall e \in [\rho],\ x \in \mathcal{A}_e \right\}.$$

This is the *marginal polytope* induced by the factorization (Wainwright and Jordan, 2008). For every $a \neq a^\star$, let $\Delta(a) = \theta(a^\star) - \theta(a) > 0$ denote its suboptimality gap, and define $\Delta_{\min} = \min_{a \neq a^\star} \Delta(a)$. We use the convention $0^{-1} = +\infty$, and let $\mathrm{kl}(p, q)$ denote the Bernoulli Kullback-Leibler divergence between two Bernoulli distributions of means p and q.

Theorem 2 (Vannella et al. (2023a))**.** *Any δ-PAC algorithm satisfies $\mathbb{E}_\theta[\tau] \geq 2\sigma^2 T_\theta^\star \, \mathrm{kl}(1 - \delta, \delta)$, where*

$$T_\theta^\star = \inf_{\tilde{w} \in \tilde{\Lambda}} \max_{a \neq a^\star} \frac{\sum_{e \in [\rho]: a_e \neq a_e^\star} \left(\tilde{w}_{e,a_e^\star}^{-1} + \tilde{w}_{e,a_e}^{-1} \right)}{\Delta(a)^2}. \tag{1}$$

The proof is reported for completeness in App. B. The *characteristic time* $T_\theta^\star$ characterizes its asymptotic difficulty. Any optimizer $\tilde{w}^\star \in \tilde{\Lambda}$ of (1) specifies the information-optimal factor-level sampling frequencies. This allocations are typically leveraged in the design of optimal BAI algorithms Garivier and Kaufmann (2016). In the following section, we show how to compute these optimal allocations by exploiting the graphical structure of the factored reward model.

5 JUNCTION-TREE REPRESENTATION OF THE CHARACTERISTIC TIME

We represent both objects in (1) on a common JT: consistent marginals encode feasible exploration, and gap-conditioned messages encode all competitors. Their combination turns the information-allocation problem into an explicit convex program.

Junction-tree setup. We use a standard JT construction associated with the reward factorization (Lauritzen and Spiegelhalter, 1988b; Wainwright and Jordan, 2008; Koller and Friedman, 2009). The scopes $\{\mathcal{S}_e\}_{e=1}^{\rho}$ define a factor graph $\mathcal{G}_f = (\mathcal{V}, \mathcal{F}, \mathcal{E}_f)$, with $\mathcal{V} = [N]$, $\mathcal{F} = [\rho]$, and $(i, e) \in \mathcal{E}_f$ iff $i \in \mathcal{S}_e$. We then define the primal graph $\mathcal{G}_p = ([N], E_p)$ as an undirected graph satisfying $(i, j) \in E_p$ iff i and j occur in a common factor scope.

For an arbitrary elimination order O of $\mathcal{G}_p$, let $\mathcal{T}_O = (\mathcal{C}_O, \mathcal{E}_O)$ denote the resulting JT. Standard JT theory guarantees that every factor scope $\mathcal{S}_e$ is contained in some bag $C \in \mathcal{C}_O$ and that the bags satisfy the running-intersection property (Lauritzen and Spiegelhalter, 1988b). Let w_O be the induced width of O, so that $|C| \leq w_O + 1$ for every $C \in \mathcal{C}_O$; consequently, each bag contains at most K^{w_O+1} local action configurations. The construction is valid for any O, while the choice of elimination order determines the induced bags and thus the computational complexity. For adjacent bags C, D, we write $S_{CD} = C \cap D$ for their separator. Figure 1 illustrates $\mathcal{G}_f$, $\mathcal{G}_p$, and the JT obtained for one elimination order. App. C gives the complete construction and the graph-theoretic details.

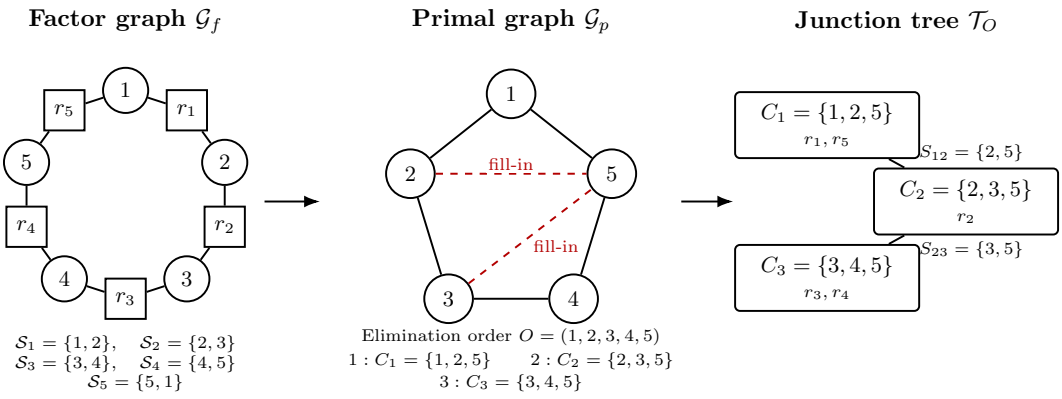

Figure 1: **From reward factors to junction tree.** The reward factor graph induces a primal graph; elimination introduces fill-in edges and bags; their inclusion-maximal elements form JT bags.

Globally feasible allocations. The first difficulty in (1) is that $\tilde{\Lambda}$ is defined through a distribution over the global action space $\mathcal{A}$. Once a JT is fixed, this global realizability constraint can be expressed exactly in terms of locally consistent distributions over its bags.

For each bag $C \in \mathcal{C}_O$, let $\mathcal{A}_C = \prod_{i \in C} \mathcal{A}_i$ and introduce a bag distribution $\mu_C \in \Delta(\mathcal{A}_C)$. For any $S \subseteq C$, let μ_C^S denote the marginal of μ_C on $\mathcal{A}_S$. We define the set of separator-consistent bag marginals as $\mathcal{M}_O \triangleq \left\{\mu = (\mu_C)_{C \in \mathcal{C}_O} : \mu_C \in \Delta(\mathcal{A}_C) \forall C \in \mathcal{C}_O,\ \mu_C^{S_{CD}} = \mu_D^{S_{CD}} \forall (C, D) \in \mathcal{E}_O\right\}$. By the running-intersection property, consistency on the JT separators is sufficient for global realizability (Wainwright and Jordan, 2008). Since every factor scope $\mathcal{S}_e$ is contained in at least one bag, fix for each $e \in [\rho]$ a host bag $C(e) \in \mathcal{C}_O$ such that $\mathcal{S}_e \subseteq C(e)$. For any $\mu \in \mathcal{M}_O$ and $x \in \mathcal{A}_e$, define $\tilde{w}_{e,x}^{\mu} \triangleq \sum_{y \in \mathcal{A}_{C(e)} : y_{\mathcal{S}_e} = x} \mu_{C(e)}(y)$. The running-intersection property together with separator consistency implies that $\tilde{w}_{e,x}^{\mu}$ is independent of the particular host bag $C(e)$ (see App D.2).

Proposition 3 (Exact representation of feasible allocations)**.** *For every JT $\mathcal{T}_O$, $\tilde{\Lambda} = \{\tilde{w}^{\mu} : \mu \in \mathcal{M}_O\}$.*

The proof is given in App. D.2. Consequently, optimization over $\tilde{\Lambda}$ can be executed using normalized bag marginals and separator-consistency constraints, with at most K^{w_O+1} configurations per bag, instead of an explicit distribution over the K^N global actions. Prop. 3 is a direct consequence of the standard JT consistency theorem. Any distribution over $\mathcal{A}$ induces separator-consistent bag marginals, while any $\mu \in \mathcal{M}_O$ admits a globally consistent distribution whose JT marginals bags are $(\mu_C)_{C \in \mathcal{C}_O}$.

Competing actions. We next address the maximization over $a \neq a^\star$ in (1). Fix $\tilde{w} \in \tilde{\Lambda}$. Define $\Delta_e(x) = \theta_e(a_e^\star) - \theta_e(x)$, set $c_{e,\tilde{w}}(a_e^\star) = 0$, and let $c_{e,\tilde{w}}(x) = \tilde{w}_{e,a_e^\star}^{-1} + \tilde{w}_{e,x}^{-1}$ for $x \neq a_e^\star$. Both $\Delta(a) = \sum_e \Delta_e(a_e)$ and $c_{\tilde{w}}(a) = \sum_e c_{e,\tilde{w}}(a_e)$ are additive, but their ratio $c_{\tilde{w}}(a)/\Delta(a)^2$ is not. For each attainable total gap s, introduce

$$\Phi_{\theta,\tilde{w}}(s) = \max\{c_{\tilde{w}}(a) : a \in \mathcal{A},\ \Delta(a) = s\}, \quad T_\theta(\tilde{w}) = \max_{s>0} \frac{\Phi_{\theta,\tilde{w}}(s)}{s^2}.$$

Conditioning on s separates the global denominator from the additive numerator. Ordinary max-sum does not directly optimize this ratio (see App. C). Augmenting the extra coordinate retains the required gap information.

To make the computation explicit, root the JT at R. Let $\mathrm{ch}(C)$ be the children of C, S_C its parent separator, and $\mathcal{E}_C = \{e : C(e) = C\}$. Define $g_C(y) = \sum_{e\in\mathcal{E}_C} \Delta_e(y_{\mathcal{S}_e})$ and $c_C(y) = \sum_{e\in\mathcal{E}_C} c_{e,\tilde{w}}(y_{\mathcal{S}_e})$. The subtree message at separator assignment z and cumulative gap s satisfies

$$M_C(z,s) = \max_{\substack{y\in\mathcal{A}_C,\ y_{S_C}=z\,;\ (s_D)_{D\in\mathrm{ch}(C)} \\ g_C(y)+\sum_D s_D = s}} \left\{ c_C(y) + \sum_{D\in\mathrm{ch}(C)} M_D(y_{S_D}, s_D) \right\}.$$

Only attainable child states are considered; at the root $S_R = \varnothing$ and $M_R(\varnothing, s) = \Phi_{\theta,\tilde{w}}(s)$. Children are merged one at a time and intermediate gaps may be *negative*. Let $H_\theta(O)$ be the largest number of attainable gap values at any fixed bag or separator assignment in these messages *and intermediate merges*. The root, hosts, and child order are fixed auxiliary choices suppressed in this notation. Appendix D.3 gives the full recursion and boundary conditions.

Proposition 4 (Gap-augmented competitor representation)**.** *For any $\tilde{w} \in \tilde{\Lambda}$ and elimination order O, the gap-augmented JT computation recovers $\Phi_{\theta,\tilde{w}}(s)$ exactly for every attainable gap s. Its time and memory complexities are $O\big(NK^{w_O+1}H_\theta(O)^2\big)$ and $O\big(NK^{w_O+1}H_\theta(O)\big)$, respectively.*

Equivalent JT formulation. Prop. 3 provides a JT representation of the feasible allocation set, while Prop. 4 provides an equivalent representation of the maximization over competing actions. Combining these two representations yields a convex program equivalent to (1). The reciprocals $\tilde{w}_{e,x}^{-1}$ are represented via rotated second-order-cone constraints, while the maxima arising in the gap-augmented elimination are represented by linear epigraph constraints. The next proposition characterize the structure of this equivalent program, while its complete is presented in App. D.4.

Theorem 5 (JT reformulation of the characteristic time)**.** *For any instance θ and elimination order O, let w_O and $H_\theta(O)$ denote the corresponding induced and gap-state widths. The optimization defining $T_\theta^\star$ admits an equivalent SOCP with $O\left((NH_\theta(O)+\rho)K^{w_O+1}\right)$ scalar variables and $O\left((NH_\theta(O)^2+\rho)K^{w_O+1}\right)$ constraints. Its optimal value is $T_\theta^\star$, and any optimal solution yields an information-optimal factor allocation $\tilde{w}^\star$.*

The proof is given in App. D.4. Thm. 5 removes the explicit dependence on the K^N joint-action space from both the feasible allocation set and the competitor optimization. The remaining complexity is governed by the induced width w_O and the gap-state width $H_\theta(O)$.

Accuracy-complexity trade-off. The exact reformulation of Thm. 5 may still be expensive when $H_\theta(O)$ is large. If the local means lie on a common lattice $\gamma\mathbb{Z}$, however, the gap-state width is bounded as $H_\theta(O) \le 1 + \gamma^{-1}\sum_{e=1}^{\rho} B_e$, where $B_e \triangleq \max_x \theta_e(x) - \min_x \theta_e(x)$. For arbitrary real-valued instances, the same structure can be recovered approximately by quantization.

Proposition 6 (Certified $(1+\varepsilon)$ approximation)**.** *For any $\varepsilon \in (0,1]$, let $\eta_\varepsilon \triangleq \frac{\varepsilon\Delta_{\min}}{8\rho}$, and $\theta_e^\varepsilon(x) \triangleq \eta_\varepsilon \,\mathrm{round}\Big(\frac{\theta_e(x)}{\eta_\varepsilon}\Big)$. Then θ^ε has the same unique optimal action as θ. Moreover, defining $\overline{T}_{\theta,\varepsilon} \triangleq \big(1+\frac{\varepsilon}{8}\big)^2 T_{\theta^\varepsilon}^\star$, $B_e^\varepsilon \triangleq \max_x \theta_e^\varepsilon(x) - \min_x \theta_e^\varepsilon(x)$ we have*

$$T_\theta^\star \le \overline{T}_{\theta,\varepsilon} \le (1+\varepsilon)T_\theta^\star, \quad \textit{and} \quad H_{\theta^\varepsilon}(O) \le 1 + \frac{8\rho}{\varepsilon\Delta_{\min}}\sum_{e=1}^{\rho} B_e^\varepsilon.$$

Thus, for fixed instance parameters, a $(1+\varepsilon)$ certificate requires $H_{\theta^\varepsilon}(O) = O(1/\varepsilon)$. Combined with Thm. 5, the gap-dependent numbers of variables and constraints scale as $O(1/\varepsilon)$ and $O(1/\varepsilon^2)$, respectively. The proof and sharper constants are given in App. D.5.

6 JUNCTION-TREE TRACK-AND-STOP

We now turn the JT characteristic-time formulation into an adaptive identification algorithm. *Junction-Tree Track-and-Stop* (JT-TaS), presented in Alg. 1, follows the TaS principle (Garivier and Kaufmann, 2016): estimate the unknown instance, compute an information-optimal sampling target, track that target while collecting observations, and stop once the accumulated evidence separates the empirical best action from all alternatives. The JT provides a representation for these operations: empirical factor means define a target allocation through the JT characteristic-time program; consistent bag marginals are converted into valid joint actions by JT sampling; and competing alternatives are evaluated through the same gap-augmented computation used in Sec. 5. Since empirical means are real-valued, we quantize them when computing targets and refine the grid over time.

Algorithm 1 JT-TAS: Junction-Tree Track-and-Stop

Require: $O, \mathcal{T}_O, \mathcal{A}_0, \mu^0, \mathcal{U}, \mathcal{Q}, \eta_t, \beta(t,\delta), \delta$

Play each $a \in \mathcal{A}_0$ once; set $\mu \leftarrow \mu^0$ and $j \leftarrow 1$
for $t = n_0, n_0+1, \dots$ **do**
  if $t \in \mathcal{U} \cup \mathcal{Q}$ **then**
    Form $\hat{\theta}_t^{\eta_t} = Q_{\eta_t}(\hat{\theta}_t)$
  end if
  if $t \in \mathcal{U}$ and $\hat{\theta}_t^{\eta_t}$ has a unique maximizer **then**
    $\mu \leftarrow$ any $\mu' \in \arg\min_{\nu \in \mathcal{M}_O} T_{\hat{\theta}_t^{\eta_t}}(\nu)$ TARGET
  end if
  if $t \in \mathcal{Q}$ and $\hat{\theta}_t, \hat{\theta}_t^{\eta_t}$ have the same unique maximizer $\hat{a}_t$ **then**
    Compute $\underline{Z}_t$ by (3) STOP
    if $\underline{Z}_t > \beta(t,\delta)$ **then**
      return $\hat{a}_t$
    end if
  end if
  if $t+1$ is a perfect square **then**
    $a_{t+1} \leftarrow a^{(j)}$; $j \leftarrow 1 + (j \bmod n_0)$ EXPLORE
  else
    $a_{t+1} \sim$ JT-Sample(μ) TRACK
  end if
  Play a_{t+1}, observe factor rewards, and update $\hat{\theta}_{t+1}$ and $\bar{\mu}_{t+1}$
end for

Empirical JT representation. Fix the JT $\mathcal{T}_O$ and the factor hosts from Sec 5. Let $\mathcal{A}_0 \subseteq \mathcal{A}$ be a *covering set* such that, for every factor configuration (e,x) with $x \in \mathcal{A}_e$, there exists some $a \in \mathcal{A}_0$ satisfying $a_e = x$. We write $n_0 = |\mathcal{A}_0|$ and choose $\mathcal{A}_0$ with duplicate completions removed, so that $n_0 \le \widetilde{A} := \sum_{e=1}^{\rho} |\mathcal{A}_e|$. The actions in $\mathcal{A}_0$ are played once initially, ensuring $N_{n_0,e,x} \ge 1$ for every (e,x), and are later reused for forced exploration.

Estimates. For $t \ge n_0$, define the empirical bag frequencies and local reward estimates by

$$\bar{\mu}_{t,C,y} = \frac{1}{t}\sum_{s=1}^{t} \mathbf{1}\{a_{s,C} = y\}, \qquad \hat{\theta}_{t,e,x} = \frac{\sum_{s=1}^{t} r_{s,e}\mathbf{1}\{a_{s,e} = x\}}{N_{t,e,x}}.$$

Since every played action is globally feasible, $\bar{\mu}_t \in \mathcal{M}_O$. For $\mu \in \mathcal{M}_O$, write $\mu_{C,y} = \mu_C(y)$ and $\mu_{e,x} = \tilde{w}^{\mu}_{e,x}$, so that $N_{t,e,x} = t\bar{\mu}_{t,e,x}$. The empirical bag potential is $\hat{\theta}_{t,C,y} = \sum_{e \in \mathcal{E}_C} \hat{\theta}_{t,e,y_{\mathcal{S}_e}}$. Observations are pooled by factor configuration. We abbreviate $T_\theta(\mu) = T_\theta(\tilde{w}^{\mu})$ for any unique-best instance θ. In the following μ denotes a target allocation, whereas $\bar{\mu}_t$ denotes the realized allocation.

Quantized JT targets. Direct optimization of $T_{\hat{\theta}_t}$ may require many real-valued gap states even at small induced width. Set $\eta_t = c_\eta / \log(t+2)$, with $c_\eta > 0$ independent of the unknown instance and of δ, and let $Q_\eta(z) = \eta\,\mathrm{round}(z/\eta)$. Round each factor estimate *before* aggregating into bags: $\hat{\theta}^{\eta_t}_{t,e,x} = Q_{\eta_t}(\hat{\theta}_{t,e,x})$ and $\hat{\theta}^{\eta_t}_{t,C,y} = \sum_{e \in \mathcal{E}_C} \hat{\theta}^{\eta_t}_{t,e,y_{\mathcal{S}_e}}$.

At update times $\mathcal{U} = \{n_0\} \cup \{2^k : k \geq 0,\ 2^k > n_0\}$ for which the quantized instance has a unique maximizer, compute

$$\mu_u \in \arg\min_{\mu \in \mathcal{M}_O} T_{\hat{\theta}_u^{\eta_u}}(\mu), \qquad T^\star_{\hat{\theta}_u^{\eta_u}} = T_{\hat{\theta}_u^{\eta_u}}(\mu_u), \tag{2}$$

using Thm. 5. Retain the target until the next successful update, leaving it unchanged at a quantized tie. We initialize μ^0 as the distributions induced by independent uniform agent actions. The target is optimal for the *quantized empirical instance*. Unlike the offline certificate, the online grid needs neither the unknown $\Delta_{\min}$ nor a prescribed relative error at each update. Forced exploration gives $\hat{\theta}_t \to \theta$ almost surely (a.s.). With $\eta_t \to 0$, this yields $\hat{\theta}_t^{\eta_t} \to \theta$.

Lemma 7 (Consistency of the JT target)**.** *Under the exploration rule below, a.s. all sufficiently late target updates have a unique quantized maximizer and, along $u \in \mathcal{U}$, $T^\star_{\hat{\theta}_u^{\eta_u}} \to T^\star_\theta$ and $T_\theta(\mu_u) \to T^\star_\theta$.*

The proof in App. E.2 compares the objectives under vanishing parameter perturbations. It also permits feasible targets with a vanishing relative optimization error.

JT Sampling . A target $\mu_u \in \mathcal{M}_O$ is represented through locally consistent bag marginals and can be sampled exactly without constructing a distribution over K^N joint actions. Root $\mathcal{T}_O$ at an arbitrary bag R. JT-TaS samples the root assignment from $\mu_{u,R}$ and then recursively samples each child conditional on the separator assignment inherited from its parent. For a child C of P with separator $S = C \cap P$, the conditional probability of a compatible residual assignment is $\mu_{u,C,y}/\mu_{u,S,y_S}$, where $\mu_{u,S}$ is the common separator marginal. The resulting joint action has bag marginals μ_u.

Forced exploration. Forced exploration samples $\mathcal{A}_0$ to ensure convergence of the parameter estimates. This gives $N_{t,e,x} = \Omega(\sqrt{t})$ for every (e, x) with only $O(\sqrt{t}) = o(t)$ forced rounds. Local means are therefore strongly consistent without altering allocation frequencies. Random target sampling tracks the time average of the retained targets, yielding the following result.

Lemma 8 (JT allocation tracking)**.** *The empirical JT allocation satisfies $T_\theta(\bar{\mu}_t) \to T^\star_\theta$ a.s..*

The proof is repored in App. E.3 combines target consistency, convexity of T_θ, vanishing exploration, and a martingale law of large numbers.

JT stopping rule. Let $\hat{a}_t$ be an empirical best joint action, obtained by max-sum (see App. C). When it is unique, minimizing the GLR ratio over all competitors gives the exact identity $Z_t = \frac{t}{2\sigma^2 T_{\hat{\theta}_t}(\bar{\mu}_t)}$. Thus JT-TaS optimizes $T_{\hat{\theta}_u^{\eta_u}}(\mu)$ to determine *where to sample*, whereas Z_t evaluates the same functional at the *realized* allocation to determine the stopping evidence. Evaluating this identity does not require re-solving the allocation problem.

Computing Z_t exactly may again require a large real-valued gap-state space. We therefore evaluate a conservative statistic on the current quantized empirical instance. Write $\Phi_{\theta,\mu}(s) \triangleq \Phi_{\theta,\tilde{w}^\mu}(s)$ for the gap profile of Section 5. A check is active only when the empirical and quantized instances have the *same unique maximizer*; otherwise set $\underline{Z}_t = 0$. At an active check, compute

$$\underline{Z}_t = \frac{t}{2\sigma^2} \min_{\substack{s>0 \\ s \text{ attainable}}} \frac{[s - \rho\eta_t]_+^2}{\Phi_{\hat{\theta}_t^{\eta_t}, \bar{\mu}_t}(s)}, \tag{3}$$

where $[z]_+ = \max\{z, 0\}$. The profile retains the largest reciprocal-marginal numerator among competitors with the same quantized gap s. Nearest-grid rounding changes every global empirical gap by at most $\rho\eta_t$, so the subtraction in (3) prevents the computation from overstating the evidence. The proof is deferred to App. E.4.

Proposition 9 (Conservative JT evidence)**.** *At every stopping check, $0 \leq \underline{Z}_t \leq Z_t$. Almost surely, all sufficiently late checks are active and $\underline{Z}_t/Z_t \to 1$.*

Let $\beta(t, \delta)$ be the count-adaptive Gaussian GLR boundary specified from Kaufmann and Koolen (2021, Prop. 15). It satisfies $\Pr_\theta(\exists t \geq n_0 : \hat{a}_t \neq a^\star,\ Z_t > \beta(t, \delta)) \leq \delta$ and $\beta(t, \delta) = \log(1/\delta) + o(\log(1/\delta))$ at horizons of order $\log(1/\delta)$. JT-TaS evaluates the stopping recursion at square times $\mathcal{Q} = \{m^2 : m \geq 1,\ m^2 \geq n_0\}$, using current empirical and quantized tables even when the sampling target is cached. It stops at $\tau_\delta = \inf\{t \in \mathcal{Q} : \underline{Z}_t > \beta(t, \delta)\}$ and returns $\hat{a}_{\tau_\delta}$.

Asymtotic optimality. The previous results show that the empirical allocation approaches the information-optimal, while the stopping statistic accumulates evidence at the corresponding optimal rate. Combining these properties with the δ-PAC stopping boundary yields the following result.

Theorem 10 (Asymptotic optimality)**.** *JT-TaS is δ-PAC, it stops in finite time, and satisfies:*

$$\Pr_{\theta}\left(\limsup_{\delta\downarrow 0}\frac{\tau_\delta}{\log(1/\delta)}\le 2\sigma^2 T^\star_\theta\right)=1,\quad \lim_{\delta\downarrow 0}\frac{\mathbb{E}_\theta[\tau_\delta]}{\log(1/\delta)}=2\sigma^2 T^\star_\theta.$$

The theorem matches the information-theoretic lower bound. Forced exploration yields consistent empirical estimates. Lem. 7 makes the quantized JT targets asymptotically optimal. Lem. 8 transfers this optimality to the realized allocation. Prop. 9 makes the computable stopping statistic asymptotically equivalent to the exact GLR. The probability-one statement uses the common sampling process, whose action-selection rule is independent of δ. A separate concentration argument establishes the expected stopping-time limit. The complete proof is given in App. E.5.

Computational cost and implementation. For every fixed instance, the empirical reward ranges are almost surely eventually bounded. Since $\eta_t = \Theta(1/\log t)$, the gap-state width $H_t = H_{\hat{\theta}_t^{\eta_t}}(O)$ of sufficiently late quantized computations satisfies $H_t = O(\log t)$. Consequently, a stopping evaluation requires $O((N\log^2 t+\rho)K^{w_O+1})$ arithmetic operations and $O((N\log t+\rho)K^{w_O+1})$ memory, including construction of the local contributions. A target update uses an SOCP with $O((N\log t+\rho)K^{w_O+1})$ variables and $O((N\log^2 t+\rho)K^{w_O+1})$ constraints. The schedules $\mathcal{U}$ and $\mathcal{Q}$ require $O(\log T)$ target solves and $O(\sqrt{T})$ stopping evaluations up to round T. The aggregate bounds are given in App. E.6.

The algorithm operations also admit a coordinator-assisted implementation: hosts maintain local statistics, sampling propagates separator assignments, and stopping exchanges separator-gap messages. The target program is solved by the coordinator, rather than by a fully decentralized optimization procedure (see App. E.7).

7 Experiments

In this section, we present numerical results on the performance of JT-TaS. We evaluate three questions: (i) whether the JT re-formulation recovers the exact characteristic time, (ii) whether retaining allocation dependence and action-dependent gaps improves sampling efficiency, (iii) and how computation scales with the graphical decomposition. Unless stated otherwise, Rewards are Gaussian with $\sigma = 0.5$, and controlled instances are rescaled to $\Delta_{\min} = 0.35$.Results are averaged over 50 independent repetitions. Additional experimental results are presented in App. F.

JT reformulation accuracy. Fig. 2(a) compares the JT formulation with exhaustive optimization, the common-gap JT approximation, and the MF restriction used in Vannella et al. (2023a). Common-gap JT preserves the structured allocation but replaces each competitor-dependent gap by the global minimum gap $\Delta_{\min}$, while MF additionally restricts allocations to product-form distributions. JT matches the exhaustive characteristic time to numerical precision across both the controlled benchmark and the additional graph families. In contrast, the approximations can substantially inflate the target: on the grid instance, common-gap JT and MF reach respectively $10.59\times$ and $11.35\times$ the exhaustive value. These results isolate the effect of retaining action-dependent gaps from the additional restriction imposed by product-form allocations.

Computational scaling. Fig. 2(b) reports the time required to solve the structured target problem across several graphical families. On the nested binary chain, whose induced width remains one, the number of reachable gap states grows from 10 at $N = 4$ to 194 at $N = 48$. The largest instance therefore contains 2^{48} joint actions, while its target is obtained in 2.43 seconds without enumerating the joint action space. The other graph families illustrate that practical cost depends not only on N, but also on the induced width and the number of attainable gap states.

Online identification. We compare JT-TaS with P-FWS (Tzeng et al., 2023), CombGame-OFW (Jourdan et al., 2021), common-gap JT sampling, and a true-mean Oracle-JT reference under the same stopping calibration. Fig. 2(c) considers the heterogeneous-chain family with $N = 5$ while increasing the local action space from $K = 2$ to $K = 4$. JT-TaS remains the best implementable method throughout the sweep. At $K = 2$, JT-TaS stops after 75.8 samples on average compared with 99.7 for P-FWS; at $K = 4$, the corresponding means are 222.0 and 1387.6. Common-gap JT reaches 517.0 samples at $K = 4$, indicating that retaining the graphical allocation alone does not recover the improvement obtained when competitor-dependent gaps are preserved.

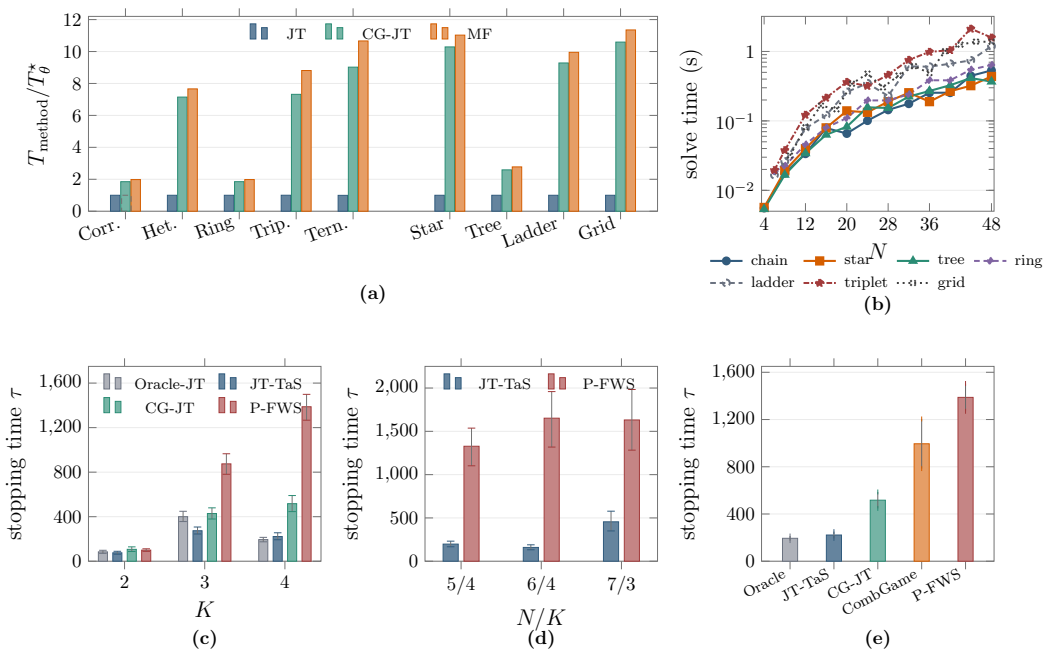

Figure 2: **Experimental results.** (a) Characteristic-time value relative to exhaustive $T^\star_\theta$ across the controlled benchmark and additional graph topologies. (b) Runtime of the structured target solver as the number of agents increases. (c) Online performance on the heterogeneous-chain family as the local action space grows. (d) Online performance on larger heterogeneous instances. (e) Comparison of the main algorithms on the heterogeneous $N = 5$, $K = 4$ instance.

Fig. 2(d) considers additional heterogeneous instances with larger local or joint action spaces. JT-TaS obtains mean stopping times of 197.9, 159.4, and 455.2, compared with 1327.9, 1651.9, and 1631.4 for P-FWS, respectively. These results show that the advantage can become particularly pronounced on heterogeneous instances with strongly non-uniform competitor gaps.

Finally, Fig. 2(e) compares all principal methods on the heterogeneous instance with $N = 5$ and $K = 4$. JT-TaS achieves a mean stopping time of 222.0, compared with 517.0 for common-gap JT, 994.7 for CombGame-OFW, and 1387.6 for P-FWS. The Oracle-JT reference reaches 194.1, showing that JT-TaS approaches the allocation obtained when the true means are known. Oracle-JT is included only as a diagnostic reference and is not an implementable learner.

8 Conclusion

In this paper, we studied fixed-confidence BAI in factored MAMABs. We showed that the information-allocation problem underlying optimal identification can be expressed through a JT, leading to a structure-aware characterization of the characteristic time and to JT-TaS, a δ-PAC and asymptotically optimal algorithm. Our results identify induced width and gap-state width as the key computational quantities governing information-optimal exploration, a picture supported by our numerical results. Potential future directions include extending our graphical approach to online learning objectives, such as regret minimization, and to settings in which the factorization of the reward function is unknown and must be learned from data.

AI Use Statement

Generative AI tools assisted with manuscript wording and organization, graphical-model preliminaries and illustrative examples, reference discovery, and feedback on notation, mathematical exposition and review, and experiments execution and presentation. The authors retain responsibility for all mathematical claims, experimental results, references, and final content.

REFERENCES

Srinivas M. Aji and Robert J. McEliece. The generalized distributive law. *IEEE Transactions on Information Theory*, 46(2):325–343, 2000. doi: 10.1109/18.825794.

Kareem Amin, Michael J. Kearns, and Umar Syed. Graphical models for bandit problems. In *Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence*, pages 1–10, 2011.

Kazuoki Azuma. Weighted sums of certain dependent random variables. *Tohoku Mathematical Journal*, 19(3):357–367, 1967. doi: 10.2748/tmj/1178243286. URL https://doi.org/10.2748/tmj/1178243286.

Eugenio Bargiacchi, Timothy Verstraeten, Diederik M. Roijers, Ann Nowé, and Hado van Hasselt. Learning to coordinate with coordination graphs in repeated single-stage multi-agent decision problems. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 482–490, 2018.

Eugenio Bargiacchi, Raphaël Avalos, Timothy Verstraeten, Pieter Libin, Ann Nowé, and Diederik M. Roijers. Multi-agent RMax for multi-agent multi-armed bandits. In *Proceedings of the Adaptive and Learning Agents Workshop at AAMAS*, pages 1–6, 2022. URL https://ala2022.github.io/papers/ALA2022_paper_38.pdf.

Alexander Bauer, Shinichi Nakajima, and Klaus-Robert Müller. Polynomial-time constrained message passing for exact MAP inference on discrete models with global dependencies. *Mathematics*, 11 (12):2628, 2023. doi: 10.3390/math11122628.

Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

Shouyuan Chen, Tian Lin, Irwin King, Michael R. Lyu, and Wei Chen. Combinatorial pure exploration of multi-armed bandits. In *Advances in Neural Information Processing Systems*, volume 27, 2014.

Thibaut Cuvelier, Richard Combes, and Eric Gourdin. Asymptotically optimal strategies for combinatorial semi-bandits in polynomial time. In *Proceedings of the 32nd International Conference on Algorithmic Learning Theory*, volume 132 of *Proceedings of Machine Learning Research*, pages 505–528, 2021.

Rina Dechter. Bucket elimination: A unifying framework for reasoning. *Artificial Intelligence*, 113 (1–2):41–85, 1999.

Aurélien Garivier and Emilie Kaufmann. Optimal best arm identification with fixed confidence. In *Proceedings of the 29th Annual Conference on Learning Theory*, volume 49 of *Proceedings of Machine Learning Research*, pages 998–1027, 2016.

Tianyuan Jin, Hao-Lun Hsu, William Chang, and Pan Xu. Finite-time frequentist regret bounds of multi-agent thompson sampling on sparse hypergraphs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 12956–12964, 2024. doi: 10.1609/aaai.v38i11.29193.

Marc Jourdan, Mojmír Mutný, Johannes Kirschner, and Andreas Krause. Efficient pure exploration for combinatorial bandits with semi-bandit feedback. In *Proceedings of the 32nd International Conference on Algorithmic Learning Theory*, volume 132 of *Proceedings of Machine Learning Research*, pages 805–849, 2021.

Emilie Kaufmann and Wouter M. Koolen. Mixture martingales revisited with applications to sequential tests and confidence intervals. *Journal of Machine Learning Research*, 22(246):1–44, 2021.

Emilie Kaufmann, Olivier Cappé, and Aurélien Garivier. On the complexity of best-arm identification in multi-armed bandit models. *Journal of Machine Learning Research*, 17(1):1–42, 2016.

Daphne Koller and Nir Friedman. *Probabilistic Graphical Models: Principles and Techniques*. MIT Press, 2009. ISBN 9780262013192.

Frank R. Kschischang, Brendan J. Frey, and Hans-Andrea Loeliger. Factor graphs and the sum-product algorithm. *IEEE Transactions on Information Theory*, 47(2):498–519, 2001. doi: 10.1109/18.910572.

Branislav Kveton, Zheng Wen, Azin Ashkan, and Csaba Szepesvári. Tight regret bounds for stochastic combinatorial semi-bandits. In *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics*, volume 38 of *Proceedings of Machine Learning Research*, pages 535–543, 2015.

Steffen L. Lauritzen and David J. Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, 50(2):157–224, 1988a.

Steffen L. Lauritzen and David J. Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, 50(2):157–194, 1988b. doi: 10.1111/j.2517-6161.1988.tb01721.x.

Shie Mannor and John N. Tsitsiklis. The sample complexity of exploration in the multi-armed bandit problem. *Journal of Machine Learning Research*, 5:623–648, 2004. URL `https://jmlr.org/papers/v5/mannor04b.html`.

Glenn R. Shafer and Prakash P. Shenoy. Probability propagation. *Annals of Mathematics and Artificial Intelligence*, 2(1–4):327–351, 1990. doi: 10.1007/BF01531015.

Ruben Stranders, Long Tran-Thanh, Francesco M. Delle Fave, Alex Rogers, and Nicholas R. Jennings. DCOPs and bandits: Exploration and exploitation in decentralised coordination. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems*, pages 289–296, 2012.

Ruo-Chun Tzeng, Po-An Wang, Alexandre Proutiere, and Chi-Jen Lu. Closing the computational-statistical gap in best arm identification for combinatorial semi-bandits. In *Advances in Neural Information Processing Systems*, volume 36, pages 18391–18403, 2023. doi: 10.52202/075280-0809.

Filippo Vannella, Alexandre Proutiere, and Jaeseong Jeong. Best arm identification in multi-agent multi-armed bandits. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 34875–34907, 2023a.

Filippo Vannella, Alexandre Proutiere, and Jaeseong Jeong. Statistical and computational trade-off in multi-agent multi-armed bandits. In *Advances in Neural Information Processing Systems*, volume 36, 2023b.

Timothy Verstraeten, Eugenio Bargiacchi, Pieter J. K. Libin, Jan Helsen, Diederik M. Roijers, and Ann Nowé. Multi-agent thompson sampling for bandit applications with sparse neighbourhood structures. *Scientific Reports*, 10(1):6728, 2020. doi: 10.1038/s41598-020-62939-3.

Martin J. Wainwright and Michael I. Jordan. Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, 1(1–2):1–305, 2008.

Po-An Wang, Ruo-Chun Tzeng, and Alexandre Proutiere. Fast pure exploration via frank-wolfe. In *Advances in Neural Information Processing Systems*, volume 34, 2021.

Tong Yu, Branislav Kveton, Zheng Wen, Ruiyi Zhang, and Ole J. Mengshoel. Graphical models meet bandits: A variational thompson sampling approach. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 10902–10912, 2020.

APPENDIX

A Table of Notation **14**

B Lower Bound Proofs **16**

B.1 Proof of Theorem 2 16

B.2 Characterization of $T_\theta^\star$ 18

C Graphical Models and Junction-Tree Preliminaries **20**

C.1 Factorizations, factor graphs, and primal graphs 20

C.2 Variable elimination, fill-in, and induced width 20

C.3 Junction trees and their construction 21

C.4 Factor assignment and the role of separators 22

C.5 Exact optimization and probabilistic marginalization 22

C.6 Consistent marginals and exact joint-action sampling 23

C.7 Computational cost and the choice of elimination order 24

C.8 Relation to the characteristic-time computation 24

C.9 Example: a three-agent chain 25

D Proofs for Junction-Tree Reformulation **26**

D.1 Auxiliary lemmas 26

D.2 Proof of Proposition 3: globally feasible allocations 27

D.3 Proof of Proposition 4: the competitor maximization 28

D.4 Proof of Theorem 5: equivalent SOCP and size 30

D.5 Proof of Proposition 6: certified approximation 32

E Proofs for Junction-Tree Track-and-Stop **34**

E.1 Auxiliary lemmas 35

E.2 Proof of Lem. 7: consistency of the JT target 38

E.3 Proof of Lem. 8: tracking the realized JT allocation 39

E.4 Proof of Prop. 9: conservative JT evidence 41

E.4.1 The adaptive Gaussian likelihood and exact GLR 41

E.4.2 Exact computation of the quantized gap profile 42

E.4.3 Proof of the stopping certificate 43

E.5 Proof of Thm. 10: correctness and asymptotic optimality 43

E.5.1 Serialization and finite-time correctness 43

E.5.2 Almost-sure evidence growth and normalized stopping time 44

E.5.3 Quantitative late-interval estimates 45

E.5.4 Expected stopping time and completion of the theorem 46

E.6 Proof of the computational-cost statement 48

E.7 Proof of the coordinator-assisted distributed-realization statement 50

F Additional experimental results **52**

F.1 Controlled instances and allocation geometry 52

F.2 Characteristic-time accuracy across graph topologies 52

A TABLE OF NOTATION

Problem setting	
N	Number of agents
$[m]$	The set $\{1, \dots, m\}$
ρ	Number of reward factors
K	Number of actions available to each agent
$\mathcal{S} = [N]$	Set of agents
$\mathcal{A}_i$	Action set of agent i
$\mathcal{A} = \times_{i \in [N]} \mathcal{A}_i$	Set of global actions, with $\lvert\mathcal{A}\rvert = A = K^N$
$\mathcal{S}_e$	Scope of factor e, with $N_e = \lvert\mathcal{S}_e\rvert$
$\mathcal{A}_e$	Local action space of factor e, with $A_e = \lvert\mathcal{A}_e\rvert = K^{N_e}$
$\widetilde{A}$	$\sum_{e \in [\rho]} A_e$
a_e	Restriction of global action a to factor scope $\mathcal{S}_e$
$\theta_e(x)$	Mean reward of local action $x \in \mathcal{A}_e$ in factor e
θ	Bandit instance $(\theta_e(x))_{e \in [\rho],\, x \in \mathcal{A}_e}$
$\theta(a)$	Expected global reward $\sum_{e \in [\rho]} \theta_e(a_e)$
σ^2	Variance of each Gaussian factor observation
$a^\star$	Unique optimal global action under θ
$\Delta(a)$	Suboptimality gap $\theta(a^\star) - \theta(a)$
$\Delta_{\min}$	$\min_{a \neq a^\star} \Delta(a)$
Θ	Set of instances with a unique optimal global action
$\mathbb{P}_\theta, \mathbb{E}_\theta$	Probability measure and expectation under instance θ
δ	Required error probability
Characteristic time and lower bound	
Λ	Simplex of probability distributions over global actions
w_a	Sampling mass assigned to global action a
$\widetilde{w}_{e,x}$	Factor marginal induced by w, $\sum_{a : a_e = x} w_a$
$\widetilde{\Lambda}$	Set of realizable factor-level sampling allocations
$\mathrm{kl}(p, q)$	KL divergence between Bernoulli distributions of means p and q
$T^\star_\theta$	Characteristic time of instance θ
$\widetilde{w}^\star$	Information-optimal factor-level sampling allocation
Junction-tree representation	
$O = (i_1, \dots, i_N)$	Elimination order
C_k	Elimination bag $\{i_k\} \cup N_O^+(i_k)$
$N_O^+(i_k)$	Neighbors of i_k remaining when i_k is eliminated
$\mathcal{T}_O = (\mathcal{C}_O, \mathcal{E}_O)$	Junction tree induced by elimination order O
S_{CD}	Separator $C \cap D$ between adjacent bags C and D
w_O	Induced width $\max_k \lvert N_O^+(i_k)\rvert$
μ_C	Probability distribution over assignments of bag C
μ_C^S	Marginal of μ_C on $S \subseteq C$
$\mathcal{M}_O$	Set of separator-consistent JT bag marginals
$C(e)$	Host bag assigned to factor e
$\widetilde{w}^\mu$	Factor allocation induced by $\mu \in \mathcal{M}_O$
$\Delta_e(x)$	Local gap $\theta_e(a_e^\star) - \theta_e(x)$
$c_{e,\widetilde{w}}(x)$	Local reciprocal-allocation contribution associated with (e, x)
$\mathcal{D}_\theta$	Set of attainable global gaps $\{\Delta(a) : a \in \mathcal{A}\}$
$\Phi_{\theta,\widetilde{w}}(s)$	Largest numerator contribution among actions with gap s
$M_{C \to P(C)}(z, s)$	Gap-augmented message from bag C to its parent
$H_\theta(O)$	Maximum gap-state width induced by θ and elimination order O
Quantization	
η	Quantization resolution
θ^η	Quantized instance with $\theta_e^\eta(x) = \eta\,\mathrm{round}(\theta_e(x)/\eta)$
q_η	Relative gap perturbation $\rho\eta/\Delta_{\min}$
$\overline{T}_{\theta,\eta}$	Certified upper approximation $(1 + q_\eta)^2 T^\star_{\theta^\eta}$

Notation for JT-TaS

$\mathcal{A}_0$	Covering set of global actions
n_0	Cardinality $\lvert\mathcal{A}_0\rvert$
a_t	Global action played at round t
$N_{t,e,x}$	Number of observations of local action x in factor e up to time t
$\hat{\theta}_{t,e,x}$	Empirical mean of local action x in factor e at time t
$\bar{\mu}_t$	Empirical JT allocation induced by the actions played up to time t
$T_\vartheta(\mu)$	Fixed-allocation characteristic-time for instance ϑ and JT allocation μ
Q_η	Quantization operator $Q_\eta(z) = \eta\,\mathrm{round}(z/\eta)$
η_t	Online quantization resolution $c_\eta/\log(t+2)$
$\hat{\theta}_t^{\eta_t}$	Quantized empirical instance at time t
$\mathcal{U}$	Set of target-update times
μ_u	JT target allocation computed at update time u
μ^0	Initial JT allocation induced by independent uniform agent actions
$\hat{a}_t$	Empirical best global action at time t
Z_t	Exact generalized likelihood-ratio statistic
$\underline{Z}_t$	Conservative quantized JT stopping statistic
$\beta(t,\delta)$	Time-dependent stopping boundary
$\mathcal{Q}$	Set of stopping-check times
τ_δ	Stopping time of JT-TaS

B Lower Bound Proofs

B.1 Proof of Theorem 2

The proof follows the classical change-of-measure argument for fixed-confidence BAI. In particular, the general lower-bound argument is the one used in Theorem 1 of Garivier and Kaufmann (2016), whose key ingredient is the transportation inequality of Lemma 1 in Kaufmann et al. (2016). Since in our setting one global action produces a vector of independent factor observations, we use the stopped log-likelihood formulation of Lemma 19 in Kaufmann et al. (2016) and explicitly factor the likelihood across factors, as in the proof of Theorem 4.1 of Vannella et al. (2023a).

Proof. Fix an instance $\theta \in \Theta$, and let $a^\star = a^\star_\theta$ denote its unique optimal global action. For every global action $b \in \mathcal{A}$, let $N_{\tau,b} = \sum_{t=1}^{\tau} \mathbf{1}\{a_t = b\}$ be the number of times b is selected before stopping, and define $w_b = \mathbb{E}_\theta[N_{\tau,b}]/\mathbb{E}_\theta[\tau]$. Since $\sum_{b\in\mathcal{A}} N_{\tau,b} = \tau$ almost surely, we have $w_b \geq 0$ and $\sum_{b\in\mathcal{A}} w_b = 1$, hence $w \in \Lambda$. Similarly, for factor e and local action $x \in \mathcal{A}_e$, define $N_{\tau,e,x} = \sum_{t=1}^{\tau} \mathbf{1}\{a_{t,e} = x\}$. Because a local configuration x is observed at factor e exactly whenever the played global action has restriction x on $\mathcal{S}_e$, $N_{\tau,e,x} = \sum_{b\in\mathcal{A}:b_e=x} N_{\tau,b}$. Therefore $\widetilde{w}_{e,x} = \mathbb{E}_\theta[N_{\tau,e,x}]/\mathbb{E}_\theta[\tau] = \sum_{b\in\mathcal{A}:b_e=x} w_b$, and consequently $\widetilde{w} \in \widetilde{\Lambda}$.

Step 1: change of measure. Let $\lambda \in \Theta$ be another bandit instance whose optimal global action differs from $a^\star$. Let $\mathcal{F}_t$ denote the sigma-field generated by the actions and observations up to round t, and let $L_t(\theta,\lambda)$ be the log-likelihood ratio between the observations generated under θ and λ up to time t. Writing p_m for the density of $\mathcal{N}(m,\sigma^2)$, conditional independence of the factor rewards gives

$$L_t(\theta,\lambda) = \sum_{s=1}^{t}\sum_{e=1}^{\rho} \log \frac{p_{\theta_e(a_{s,e})}(r_{s,e})}{p_{\lambda_e(a_{s,e})}(r_{s,e})}. \tag{4}$$

Equivalently, $L_t(\theta,\lambda) = \sum_{s=1}^{t}\sum_{e=1}^{\rho}\sum_{x\in\mathcal{A}_e} \mathbf{1}\{a_{s,e}=x\} \log \frac{p_{\theta_e(x)}(r_{s,e})}{p_{\lambda_e(x)}(r_{s,e})}$.

Consider the event $\mathcal{E} = \{\hat{a}_\tau = a^\star\} \in \mathcal{F}_\tau$. Since the algorithm is δ-PAC, $\mathbb{P}_\theta(\mathcal{E}) \geq 1-\delta$. Moreover, because $a^\star_\lambda \neq a^\star$, $\mathbb{P}_\lambda(\mathcal{E}) \leq \delta$. Lemma 19 of Kaufmann et al. (2016), applied to the almost surely finite stopping time τ and the event $\mathcal{E}$, yields

$$\mathbb{E}_\theta[L_\tau(\theta,\lambda)] \geq \mathrm{kl}(\mathbb{P}_\theta(\mathcal{E}), \mathbb{P}_\lambda(\mathcal{E}))\,. \tag{5}$$

Since $\delta < 1/2$, we have $\mathbb{P}_\theta(\mathcal{E}) > \mathbb{P}_\lambda(\mathcal{E})$. The Bernoulli relative entropy $\mathrm{kl}(p,q)$ is increasing in p and decreasing in q whenever $p > q$. Hence

$$\mathbb{E}_\theta[L_\tau(\theta,\lambda)] \geq \mathrm{kl}(1-\delta,\delta). \tag{6}$$

Step 2: expected log-likelihood ratio under semi-bandit feedback. We now express the left-hand side of (6) in terms of the expected factor counts. Using (4) at the stopping time, $L_\tau(\theta,\lambda) = \sum_{s\geq 1}\mathbf{1}\{\tau \geq s\}\sum_{e=1}^{\rho}\sum_{x\in\mathcal{A}_e}\mathbf{1}\{a_{s,e}=x\}\log\frac{p_{\theta_e(x)}(r_{s,e})}{p_{\lambda_e(x)}(r_{s,e})}$. For every s, both $\mathbf{1}\{\tau \geq s\} = \mathbf{1}\{\tau > s-1\}$ and a_s are $\mathcal{F}_{s-1}$-measurable. Conditional on $\mathcal{F}_{s-1}$ and on $a_{s,e} = x$, the observation $r_{s,e}$ has distribution $\mathcal{N}(\theta_e(x),\sigma^2)$ under θ. Hence $\mathbb{E}_\theta\left[\log\frac{p_{\theta_e(x)}(r_{s,e})}{p_{\lambda_e(x)}(r_{s,e})}\middle|\mathcal{F}_{s-1},\, a_{s,e}=x\right] = \mathrm{KL}\big(\mathcal{N}(\theta_e(x),\sigma^2),\mathcal{N}(\lambda_e(x),\sigma^2)\big)$. Because $\mathbb{E}_\theta[\tau] < \infty$ and all Gaussian log-likelihood increments are integrable, expectation and summation can be exchanged. We obtain

$$\mathbb{E}_\theta[L_\tau(\theta,\lambda)] = \sum_{e=1}^{\rho}\sum_{x\in\mathcal{A}_e}\mathbb{E}_\theta[N_{\tau,e,x}]\,\mathrm{KL}\big(\mathcal{N}(\theta_e(x),\sigma^2),\mathcal{N}(\lambda_e(x),\sigma^2)\big)\,. \tag{7}$$

For Gaussians with common known variance, $\mathrm{KL}(\mathcal{N}(m,\sigma^2),\mathcal{N}(m',\sigma^2)) = (m-m')^2/(2\sigma^2)$. Combining this identity with (6) and (7) gives, for every alternative λ whose optimal action differs from $a^\star$,

$$\frac{\mathbb{E}_\theta[\tau]}{2\sigma^2}\sum_{e=1}^{\rho}\sum_{x\in\mathcal{A}_e}\widetilde{w}_{e,x}\big(\theta_e(x)-\lambda_e(x)\big)^2 \geq \mathrm{kl}(1-\delta,\delta). \tag{8}$$

Step 3: reduction to a fixed competing action. Fix a global action $a \neq a^\star$, and define $\mathcal{B}_a(\theta) = \{\lambda \in \Theta : \lambda(a) > \lambda(a^\star)\}$. Every $\lambda \in \mathcal{B}_a(\theta)$ is a valid confusing instance, since $a^\star$ cannot remain optimal when $\lambda(a) > \lambda(a^\star)$. Therefore (8) holds for every $\lambda \in \mathcal{B}_a(\theta)$.

Let $Q_{\widetilde{w}}(\lambda) = \sum_{e=1}^{\rho} \sum_{x \in \mathcal{A}_e} \widetilde{w}_{e,x} \big(\theta_e(x) - \lambda_e(x)\big)^2$. Taking the infimum over $\mathcal{B}_a(\theta)$ yields

$$\frac{\mathbb{E}_\theta[\tau]}{2\sigma^2} \inf_{\lambda \in \mathcal{B}_a(\theta)} Q_{\widetilde{w}}(\lambda) \geq \mathrm{kl}(1-\delta, \delta). \tag{9}$$

The strict inequality in the definition of $\mathcal{B}_a(\theta)$ can be replaced by its boundary when computing this infimum. Indeed, $Q_{\widetilde{w}}$ is continuous in λ, and every parameter satisfying $\lambda(a) = \lambda(a^\star)$ can be approached arbitrarily closely by parameters for which $\lambda(a) > \lambda(a^\star)$. If necessary, an arbitrarily small generic perturbation can simultaneously break any remaining ties between global actions, so that the approximating parameters belong to Θ. Consequently,

$$\inf_{\lambda \in \mathcal{B}_a(\theta)} Q_{\widetilde{w}}(\lambda) = \inf_{\lambda : \lambda(a) \geq \lambda(a^\star)} Q_{\widetilde{w}}(\lambda). \tag{10}$$

Let $I(a) = \{e \in [\rho] : a_e \neq a_e^\star\}$. Since $\lambda(a) - \lambda(a^\star) = \sum_{e \in I(a)} \left(\lambda_e(a_e) - \lambda_e(a_e^\star)\right)$, the constraint in (10) depends only on the two coordinates a_e and $a_e^\star$ for factors $e \in I(a)$. Every other coordinate appears only through a nonnegative quadratic term in the objective and can therefore be left unchanged, namely $\lambda_e(x) = \theta_e(x)$. For every $e \in I(a)$, define $u_e = \lambda_e(a_e) - \theta_e(a_e)$ and $v_e = \lambda_e(a_e^\star) - \theta_e(a_e^\star)$. Since $\Delta(a) = \theta(a^\star) - \theta(a) = \sum_{e \in I(a)} \left(\theta_e(a_e^\star) - \theta_e(a_e)\right)$, we can rewrite $\lambda(a) \geq \lambda(a^\star)$ as

$$\sum_{e \in I(a)} (u_e - v_e) \geq \Delta(a). \tag{11}$$

Hence

$$\inf_{\lambda : \lambda(a) \geq \lambda(a^\star)} Q_{\widetilde{w}}(\lambda) = \inf_{\sum_{e \in I(a)} (u_e - v_e) \geq \Delta(a)} \sum_{e \in I(a)} \left(\widetilde{w}_{e,a_e} u_e^2 + \widetilde{w}_{e,a_e^\star} v_e^2\right). \tag{12}$$

Step 4: solution of the quadratic program. Assume first that $\widetilde{w}_{e,a_e} > 0$ and $\widetilde{w}_{e,a_e^\star} > 0$ for every $e \in I(a)$. Set $\alpha_e = \widetilde{w}_{e,a_e}$ and $\beta_e = \widetilde{w}_{e,a_e^\star}$. Because $\Delta(a) > 0$, the unconstrained minimizer $u_e = v_e = 0$ is infeasible, and the constraint in (11) must therefore be active at the optimum.

Consider the Lagrangian $\mathcal{L}(u, v, c) = \sum_{e \in I(a)} (\alpha_e u_e^2 + \beta_e v_e^2) + 2c \left(\Delta(a) - \sum_{e \in I(a)} (u_e - v_e)\right)$, with multiplier $c \geq 0$. The stationarity conditions are $2\alpha_e u_e - 2c = 0$ and $2\beta_e v_e + 2c = 0$, so that $u_e = c/\alpha_e$ and $v_e = -c/\beta_e$. Imposing the active constraint gives $c \sum_{e \in I(a)} \left(\alpha_e^{-1} + \beta_e^{-1}\right) = \Delta(a)$. Thus $c = \Delta(a) \left[\sum_{e \in I(a)} \left(\alpha_e^{-1} + \beta_e^{-1}\right)\right]^{-1}$. Substituting the optimizer into the objective yields

$$\inf_{\lambda : \lambda(a) \geq \lambda(a^\star)} Q_{\widetilde{w}}(\lambda) = \frac{\Delta(a)^2}{\displaystyle\sum_{e \in I(a)} \left(\widetilde{w}_{e,a_e}^{-1} + \widetilde{w}_{e,a_e^\star}^{-1}\right)}. \tag{13}$$

The zero-allocation case is consistent with the same expression under the convention $0^{-1} = +\infty$. Indeed, if for some $e \in I(a)$ either $\widetilde{w}_{e,a_e} = 0$ or $\widetilde{w}_{e,a_e^\star} = 0$, the constraint (11) can be satisfied by modifying only that unobserved coordinate, at zero quadratic cost. Hence the infimum in (12) is zero, while the denominator in (13) is $+\infty$.

Define $V_{\widetilde{w}}(a) = \sum_{e \in I(a)} \left(\widetilde{w}_{e,a_e}^{-1} + \widetilde{w}_{e,a_e^\star}^{-1}\right)$. Equations (9)–(13) therefore give $\mathbb{E}_\theta[\tau]\Delta(a)^2/(2\sigma^2 V_{\widetilde{w}}(a)) \geq \mathrm{kl}(1-\delta, \delta)$. Equivalently, for every $a \neq a^\star$,

$$\mathbb{E}_\theta[\tau] \geq 2\sigma^2 \, \mathrm{kl}(1-\delta, \delta) \frac{V_{\widetilde{w}}(a)}{\Delta(a)^2}. \tag{14}$$

If $V_{\widetilde{w}}(a) = +\infty$, the preceding change-of-measure argument shows that no δ-PAC algorithm with finite expected stopping time can induce such an allocation; thus (14) remains valid in the extended-real sense.

Step 5: optimization over the sampling allocation. Since (14) holds simultaneously for all competitors, $\mathbb{E}_\theta[\tau] \geq 2\sigma^2 \operatorname{kl}(1-\delta,\delta) \max_{a\neq a^\star} \frac{V_{\widetilde{w}}(a)}{\Delta(a)^2}$. The allocation $\widetilde{w}$ appearing here is the factor marginal allocation actually induced by the algorithm, and we proved at the beginning that $\widetilde{w} \in \widetilde{\Lambda}$. Therefore

$$\begin{aligned}\mathbb{E}_\theta[\tau] &\geq 2\sigma^2 \operatorname{kl}(1-\delta,\delta) \max_{a\neq a^\star} \frac{\displaystyle\sum_{e:a_e\neq a_e^\star} \left(\widetilde{w}_{e,a_e}^{-1} + \widetilde{w}_{e,a_e^\star}^{-1}\right)}{\Delta(a)^2} \\ &\geq 2\sigma^2 \operatorname{kl}(1-\delta,\delta) \inf_{\widetilde{v}\in\widetilde{\Lambda}} \max_{a\neq a^\star} \frac{\displaystyle\sum_{e:a_e\neq a_e^\star} \left(\widetilde{v}_{e,a_e}^{-1} + \widetilde{v}_{e,a_e^\star}^{-1}\right)}{\Delta(a)^2}.\end{aligned}$$

The last infimum is exactly $T_\theta^\star$ in (1). Hence $\mathbb{E}_\theta[\tau] \geq 2\sigma^2 T_\theta^\star \operatorname{kl}(1-\delta,\delta)$, which proves Thm. 2. □

B.2 Characterization of $T_\theta^\star$

The expression of $T_\theta^\star$ in Theorem 2 is exact, but its dependence on the instance is implicit. We now provide a simpler characterization revealing the relevant statistical scaling.

For $a \neq a^\star$, let $I(a) = \{e \in [\rho] : a_e \neq a_e^\star\}$ and $s(a) = |I(a)|$, and define

$$\mathcal{H}(\theta) \triangleq \max_{a\neq a^\star} \frac{s(a)}{\Delta(a)^2}.$$

Let $m_e = |\mathcal{A}_e|$, $\widetilde{A} = \sum_{e=1}^{\rho} m_e$, and define the weighted complexity

$$\mathcal{H}_{\mathrm{w}}(\theta) \triangleq \max_{a\neq a^\star} \frac{\sum_{e\in I(a)} m_e}{\Delta(a)^2}.$$

Finally, for $q \in [0,1)$, let

$$c_m(q) = \frac{m}{1+(m-1)q} + \frac{m}{1-q}, \qquad \mathcal{U}(\theta) = \inf_{q\in[0,1)} \max_{a\neq a^\star} \frac{\sum_{e\in I(a)} c_{m_e}(q)}{\Delta(a)^2}.$$

Lemma 11 (Factor-disagreement characterization)**.** *Let* $m_{\max} = \max_{e\in[\rho]} m_e$ *and* $\Delta_{\min} = \min_{a\neq a^\star} \Delta(a)$. *Then*

$$4\mathcal{H}(\theta) \leq T_\theta^\star \leq \mathcal{U}(\theta) \leq 2\mathcal{H}_{\mathrm{w}}(\theta) \leq \frac{2\widetilde{A}}{\Delta_{\min}^2}.$$

Moreover,

$$T_\theta^\star \leq \left(1+\sqrt{m_{\max}-1}\right)^2 \mathcal{H}(\theta).$$

The constants in

$$4\mathcal{H}(\theta) \leq T_\theta^\star \leq \left(1+\sqrt{m_{\max}-1}\right)^2 \mathcal{H}(\theta)$$

are sharp in general.

Proof. Recall that

$$T_\theta^\star = \inf_{\widetilde{w}\in\widetilde{\Lambda}} \max_{a\neq a^\star} \frac{V_{\widetilde{w}}(a)}{\Delta(a)^2}, \qquad V_{\widetilde{w}}(a) = \sum_{e\in I(a)} \left(\widetilde{w}_{e,a_e}^{-1} + \widetilde{w}_{e,a_e^\star}^{-1}\right).$$

For the lower bound, fix $\widetilde{w} \in \widetilde{\Lambda}$ and $a \neq a^\star$. For every $e \in I(a)$, the configurations a_e and $a_e^\star$ are distinct and $\widetilde{w}_{e,a_e} + \widetilde{w}_{e,a_e^\star} \leq 1$. Hence, with the convention $0^{-1} = +\infty$,

$$\widetilde{w}_{e,a_e}^{-1} + \widetilde{w}_{e,a_e^\star}^{-1} \geq 4,$$

since $p^{-1} + r^{-1} \geq 4/(p+r)$ for $p, r > 0$. Therefore $V_{\widetilde{w}}(a) \geq 4s(a)$, and thus

$$T_\theta^\star \geq 4 \max_{a\neq a^\star} \frac{s(a)}{\Delta(a)^2} = 4\mathcal{H}(\theta).$$

For the upper bound, fix $q \in [0,1)$ and consider the feasible global allocation

$$w_b^{(q)} = q\,\mathbf{1}\{b = a^\star\} + \frac{1-q}{|\mathcal{A}|}, \qquad b \in \mathcal{A}.$$

Its factor marginals satisfy

$$\widetilde{w}_{e,a_e^\star}^{(q)} = \frac{1+(m_e-1)q}{m_e}, \qquad \widetilde{w}_{e,x}^{(q)} = \frac{1-q}{m_e}, \quad x \neq a_e^\star.$$

Hence, for every $a \neq a^\star$,

$$V_{\widetilde{w}^{(q)}}(a) = \sum_{e \in I(a)} \left(\frac{m_e}{1+(m_e-1)q} + \frac{m_e}{1-q} \right) = \sum_{e \in I(a)} c_{m_e}(q).$$

By feasibility of $\widetilde{w}^{(q)}$ and then optimizing over q,

$$T_\theta^\star \leq \mathcal{U}(\theta).$$

Evaluating the latter at $q = 0$, for which $c_m(0) = 2m$, gives

$$\mathcal{U}(\theta) \leq 2 \max_{a \neq a^\star} \frac{\sum_{e \in I(a)} m_e}{\Delta(a)^2} = 2\mathcal{H}_{\mathrm{w}}(\theta).$$

Since $I(a) \subseteq [\rho]$ and $\Delta(a) \geq \Delta_{\min}$,

$$\mathcal{H}_{\mathrm{w}}(\theta) \leq \frac{\widetilde{A}}{\Delta_{\min}^2},$$

which proves the first chain of inequalities.

It remains to derive the simpler factor-disagreement upper bound. For fixed q, $c_m(q)$ is increasing in m, since

$$\frac{\partial}{\partial m} \frac{m}{1+(m-1)q} = \frac{1-q}{(1+(m-1)q)^2} > 0,$$

and $m/(1-q)$ is also increasing. Thus

$$\mathcal{U}(\theta) \leq \mathcal{H}(\theta) \inf_{q \in [0,1)} c_{m_{\max}}(q).$$

A direct minimization gives

$$\inf_{q \in [0,1)} c_m(q) = \left(1 + \sqrt{m-1}\right)^2.$$

Indeed, $c_m'(q) = 0$ is equivalent to $1+(m-1)q = \sqrt{m-1}(1-q)$, whose solution attains the value above. Therefore

$$T_\theta^\star \leq \left(1 + \sqrt{m_{\max}-1}\right)^2 \mathcal{H}(\theta).$$

To see that the constants are sharp, consider first a single binary factor with gap Δ. Then $\mathcal{H}(\theta) = \Delta^{-2}$ and

$$T_\theta^\star = \frac{1}{\Delta^2} \inf_{p \in (0,1)} \left(\frac{1}{p} + \frac{1}{1-p} \right) = \frac{4}{\Delta^2}.$$

For the upper constant, consider a single factor with m actions, one optimal action and $m-1$ suboptimal actions with common gap Δ. By symmetry,

$$T_\theta^\star = \frac{1}{\Delta^2} \inf_{p \in (0,1)} \left(\frac{1}{p} + \frac{m-1}{1-p} \right) = \frac{(1+\sqrt{m-1})^2}{\Delta^2},$$

which proves sharpness. □

Remark 12 (Comparison with prior work)**.** *The refined bound preserves both the alternative-specific gaps and the factors actually modified by each competitor:*

$$T_\theta^\star \leq \mathcal{U}(\theta) \leq 2\mathcal{H}_{\mathrm{w}}(\theta) \leq \frac{2\widetilde{A}}{\Delta_{\min}^2}.$$

Hence it is never larger than the worst-case bound $2\widetilde{A}/\Delta_{\min}^2$ in Lemma 5.1 of Vannella et al. (2023a), and can be substantially smaller when difficult alternatives differ from $a^\star$ only locally. Moreover, when $m_{\max} = O(1)$,

$$T_\theta^\star = \Theta\left(\max_{a \neq a^\star} \frac{s(a)}{\Delta(a)^2} \right),$$

with constants independent of ρ and of the exponentially large global action space.

C Graphical Models and Junction-Tree Preliminaries

This appendix collects the graphical-model background used in Section 5 and fixes the conventions underlying our junction-tree (JT) representation. We describe how the reward scopes determine an elimination structure, how separators support exact local computations, and how consistent bag marginals represent globally feasible sampling allocations. These are standard graphical-model constructions (Lauritzen and Spiegelhalter, 1988b; Dechter, 1999; Koller and Friedman, 2009). Their specialization to the characteristic-time program, including the gap-augmented recursion and its conic formulation, is proved in Appendix D.

C.1 Factorizations, factor graphs, and primal graphs

For any $C \subseteq [N]$, let $\mathcal{A}_C = \prod_{i \in C} \mathcal{A}_i$, and write a_C for the restriction of $a \in \mathcal{A}$ to C. In particular, $\mathcal{A}_e = \mathcal{A}_{\mathcal{S}_e}$ and $a_e = a_{\mathcal{S}_e}$. The empty set has a single assignment, so $|\mathcal{A}_\varnothing| = 1$; sums and products over empty index sets are zero and one, respectively.

Factor graph. The decomposition $\theta(a) = \sum_{e=1}^{\rho} \theta_e(a_e)$ is represented by the bipartite graph $\mathcal{G}_f = (\mathcal{V}, \mathcal{F}, \mathcal{E}_f)$, where the two vertex classes are the agent variables $\mathcal{V} = [N]$ and the reward factors $\mathcal{F} = [\rho]$, with $(i, e) \in \mathcal{E}_f$ if and only if $i \in \mathcal{S}_e$. A factor node therefore records which action coordinates enter one local reward table. This is the usual factor-graph representation of a function decomposed into local terms (Kschischang et al., 2001). Distinct factors remain distinct nodes even when their scopes coincide.

Primal graph. The associated primal graph $\mathcal{G}_p = ([N], E_p)$ has an undirected edge $\{i, j\}$, for $i \neq j$, precisely when some factor contains both agents:

$$E_p = \bigcup_{e=1}^{\rho} \big\{\{i, j\} : i, j \in \mathcal{S}_e,\ i \neq j\big\}.$$

Thus every factor scope forms a clique, meaning a set of pairwise adjacent vertices. The primal graph retains the interactions relevant to elimination, but not the identities or multiplicities of the original factors. In particular, a clique in $\mathcal{G}_p$ need not itself be a factor scope: its edges may come from different factors.

Relation to probabilistic graphical models. The same scope structure can represent a probability mass function $p_\psi(a) = Z_\psi^{-1} \prod_e \psi_e(a_e)$, where $\psi_e \geq 0$ and $0 < Z_\psi = \sum_a \prod_e \psi_e(a_e) < \infty$. Such local functions are called *potentials*; they need not be normalized individually (Koller and Friedman, 2009). Additive optimization and probabilistic marginalization exploit the same decomposition, but use different operations when combining and eliminating terms (Aji and McEliece, 2000). In our model, the reward factorization does not require the sampling distribution $w \in \Lambda$ to factor across agents or reward scopes. Likewise, the assumed independence of reward observations conditional on the selected action is a separate statistical assumption, not a restriction on correlations among the selected actions.

C.2 Variable elimination, fill-in, and induced width

Fix an elimination order $O = (i_1, \ldots, i_N)$, with variables eliminated from left to right. Immediately before eliminating i_k, let $N_O^+(i_k)$ be its neighbors among the remaining variables, including edges added at earlier steps. Elimination first connects every pair of vertices in $N_O^+(i_k)$ and then removes i_k. The newly introduced edges are called *fill-in edges*, and $C_k = \{i_k\} \cup N_O^+(i_k)$ is the corresponding *elimination bag*. This graph operation tracks the scope of an intermediate table: after combining the terms involving i_k, eliminating that variable can leave a function of all its remaining neighbors (Dechter, 1999).

Let $\mathcal{G}_O^+$ be the graph on the original vertex set containing E_p and every fill-in edge generated by O. A graph is *chordal*, or triangulated, if every cycle of length at least four has an edge between two nonconsecutive vertices. The graph $\mathcal{G}_O^+$ is chordal: O is a perfect elimination order, since the later neighbors of every vertex form a clique. Its maximal cliques are exactly the inclusion-maximal

members of $\{C_1, \ldots, C_N\}$. We denote this collection by $\mathcal{C}_O$ and write $b_O = |\mathcal{C}_O| \leq N$. Here and below, *maximal* means inclusion-maximal, not necessarily of maximum cardinality.

The *induced width* of the chosen order and the *treewidth* of the primal graph are, respectively,

$$w_O = \max_{k \in [N]} |N_O^+(i_k)| = \max_{C \in \mathcal{C}_O} (|C| - 1), \qquad \mathrm{tw}(\mathcal{G}_p) = \min_O w_O.$$

We use the convention that width is bag size minus one. Consequently, under $|\mathcal{A}_i| = K$, a bag contains at most K^{w_O+1} assignments. For heterogeneous action sets, its exact table size is $|\mathcal{A}_C| = \prod_{i \in C} |\mathcal{A}_i|$. All our representations are valid for any chosen O; they do not require an order attaining the treewidth.

Fill-in is computational, not statistical. Triangulation groups variables for elimination. It does not add reward factors, alter θ, or change the action space. Even when every original factor is pairwise, intermediate bags can contain more than two agents. Factor arity and induced width therefore control different objects.

C.3 Junction trees and their construction

A JT for the reward scopes is a tree of bags that covers $[N]$, contains each $\mathcal{S}_e$ in at least one bag, and satisfies the *running-intersection property*: for every agent i, the bags containing i form a connected subtree. Equivalently, for any bags C, D, every bag on their unique connecting path contains $C \cap D$. In the construction used here, the bags are $\mathcal{C}_O$, the maximal cliques of $\mathcal{G}_O^+$, and the tree is denoted by $\mathcal{T}_O = (\mathcal{C}_O, \mathcal{E}_O)$. Such a tree is also called a *clique tree*; its existence follows from chordality (Lauritzen and Spiegelhalter, 1988b). More generally, bags satisfying coverage and running intersection form a tree decomposition and need not be maximal cliques of the original primal graph.

Construction from the elimination bags. One can build a JT directly from the chosen order. Whenever $N_O^+(i_k) \neq \varnothing$, connect C_k to $C_{p(k)}$, where $p(k) = \min\{\ell : i_\ell \in N_O^+(i_k)\}$. Since $p(k) > k$ and elimination makes the remaining neighbors a clique,

$$N_O^+(i_k) \subseteq C_{p(k)}, \qquad C_k \cap C_{p(k)} = N_O^+(i_k).$$

These edges form a running-intersection forest. A nonmaximal bag can be removed by contracting an edge to a neighboring bag containing it; repeating this operation leaves the bags in $\mathcal{C}_O$ without enlarging any bag. Different forest components have disjoint variable sets and may be joined by arbitrary tree edges with empty intersections. Lemma 13 gives the proof, including factor-scope containment and the disconnected case.

An equivalent construction forms the complete graph on $\mathcal{C}_O$, assigns edge CD weight $|C \cap D|$, and takes a maximum-weight spanning tree. The clique-tree theorem guarantees running intersection; zero-weight edges connect disconnected components when necessary. An arbitrary spanning tree need not have this property.

For each JT edge CD, the intersection $S_{CD} = C \cap D$ is its *separator*. Distinct maximal bags do not contain one another, so

$$|C| \leq w_O + 1, \qquad |S_{CD}| \leq w_O, \qquad |\mathcal{A}_{S_{CD}}| \leq K^{w_O}.$$

The last two bounds use the maximal-bag construction; a separator is not smaller merely because it is an intersection of two arbitrary bags.

Example: a four-agent cycle. Consider scopes $\{1,2\}, \{2,3\}, \{3,4\}, \{4,1\}$ and the order $O = (1,2,3,4)$. Eliminating agent 1 adds the edge $\{2,4\}$. The elimination bags are $\{1,2,4\}$, $\{2,3,4\}$, $\{3,4\}$, and $\{4\}$. The JT therefore has two maximal bags, $\{1,2,4\}$ and $\{2,3,4\}$, joined by separator $\{2,4\}$, with $w_O = 2$. All four original factors are covered, but no reward factor on $\{2,4\}$ has been introduced. Computation uses bag tables of size K^3 and a separator table of size K^2, rather than a table indexed by all K^4 joint actions.

C.4 Factor assignment and the role of separators

Choose exactly one *host bag* $C(e) \supseteq \mathcal{S}_e$ for each factor and let $\mathcal{E}_C = \{e : C(e) = C\}$. The sets $\mathcal{E}_C$ partition $[\rho]$. For any additive objective $f(a) = \sum_e f_e(a_e)$, define

$$f_C(y) = \sum_{e \in \mathcal{E}_C} f_e(y_{\mathcal{S}_e}), \qquad f(a) = \sum_{C \in \mathcal{C}_O} f_C(a_C).$$

A bag with no hosted factors has $f_C \equiv 0$. Assigning a factor once, rather than to every bag containing its scope, avoids counting its contribution multiple times. In particular, $f_e = \theta_e$ gives the reward potentials used in Section 6.

Root the JT at a bag R. For $C \neq R$, write $P(C)$ for its parent, $S_C = C \cap P(C)$ for its parent separator, and $\mathrm{ch}(C)$ for its children. Let U_C be the union of the bags in the subtree rooted at C, and set $S_R = \varnothing$. Removing edge $CP(C)$ splits the tree into two parts. Running intersection implies that the variables appearing on both sides are exactly S_C. Moreover, distinct child subtrees have no shared variables outside C.

Thus, after fixing a bag assignment $y \in \mathcal{A}_C$, assignments generated in different child subtrees can be combined whenever they agree with y on the corresponding separators. No additional compatibility check is required between their remaining coordinates. These statements are formalized in Lemma 14. They explain why a message needs to retain only the assignment on the separator, rather than the entire partial joint action.

The sets R and $C \setminus S_C$, for $C \neq R$, partition $[N]$: each variable is introduced at its highest bag in the rooted tree. This observation also permits a global action to be reconstructed or sampled by a parent-before-child traversal without assigning any coordinate twice.

C.5 Exact optimization and probabilistic marginalization

Max-sum messages. For the additive objective above, the message from C to its parent is a table indexed by $z \in \mathcal{A}_{S_C}$:

$$m_{C \to P(C)}(z) = \max_{\substack{y \in \mathcal{A}_C \\ y_{S_C} = z}} \left\{ f_C(y) + \sum_{D \in \mathrm{ch}(C)} m_{D \to C}(y_{S_D}) \right\}.$$

At a leaf, the sum is empty. Computing the messages from leaves to root gives

$$\max_{a \in \mathcal{A}} f(a) = \max_{y \in \mathcal{A}_R} \left\{ f_R(y) + \sum_{D \in \mathrm{ch}(R)} m_{D \to R}(y_{S_D}) \right\}.$$

The message represents the maximum total contribution of factors hosted in the subtree of C, conditional on its boundary assignment z. Subtree separation makes the child optimizations compatible, yielding the usual variable-elimination recursion (Dechter, 1999). Recording a maximizing bag assignment for each separator state permits an optimizing global action to be recovered by backtracking from the root. When scores tie, a fixed tie-breaking rule and these conditional backpointers give a consistent maximizer; independently maximizing individual bag tables does not in general do so. Taking $f_e = \theta_e$ or $f_e = \hat{\theta}_{t,e}$ computes an optimal or empirical-optimal joint action, respectively.

Sum-product messages. For nonnegative factors ψ_e, assign each factor to the same host and set $\psi_C(y) = \prod_{e \in \mathcal{E}_C} \psi_e(y_{\mathcal{S}_e})$. The corresponding message replaces maximization by summation and addition by multiplication:

$$m^{\mathrm{sp}}_{C \to P(C)}(z) = \sum_{\substack{y \in \mathcal{A}_C \\ y_{S_C} = z}} \psi_C(y) \prod_{D \in \mathrm{ch}(C)} m^{\mathrm{sp}}_{D \to C}(y_{S_D}).$$

The root sum gives Z_ψ. Passing messages in both directions gives, for every bag,

$$p_\psi(a_C = y) = \frac{1}{Z_\psi} \psi_C(y) \prod_{D:\, CD \in \mathcal{E}_O} m^{\mathrm{sp}}_{D \to C}(y_{C \cap D}),$$

where the messages are unnormalized. This is exact JT marginalization (Shafer and Shenoy, 1990). Its exactness depends on the running-intersection structure, not merely on passing messages along some tree of factor scopes. Max-sum and sum-product are instances of the same distributive computation (Aji and McEliece, 2000). Neither requires enumerating a distribution on $\mathcal{A}$.

C.6 Consistent marginals and exact joint-action sampling

A bag potential is a local function used in an objective or factorization; a bag marginal is a normalized probability distribution. These are different objects. In the allocation representation, the unknowns are marginals $\mu_C \in \Delta(\mathcal{A}_C)$, not reward potentials. For $S \subseteq C$, write

$$\mu_C^S(z) = \sum_{\substack{y \in \mathcal{A}_C \\ y_S = z}} \mu_C(y), \qquad z \in \mathcal{A}_S.$$

The separator-consistent marginal polytope is

$$\mathcal{M}_O = \left\{ (\mu_C)_{C \in \mathcal{C}_O} : \mu_C \in \Delta(\mathcal{A}_C), \quad \mu_C^{S_{CD}} = \mu_D^{S_{CD}} \text{ for every } CD \in \mathcal{E}_O \right\}.$$

JT consistency states that these local constraints are necessary and sufficient for the existence of a global distribution with the prescribed bag marginals (Wainwright and Jordan, 2008, Section 2.5.2). In particular, the factor marginal $\tilde{w}^\mu_{e,x} = \mu^{\mathcal{S}_e}_{C(e)}(x)$ is independent of the host choice: every bag and separator on a path between two possible hosts contains $\mathcal{S}_e$. Proposition 3 therefore gives the exact projection $\tilde{\Lambda} = \{\tilde{w}^\mu : \mu \in \mathcal{M}_O\}$. The proof is in Appendix D.2.

Reconstruction and sampling. Let μ_{S_C} be the common marginal on separator S_C. For each $z \in \mathcal{A}_{S_C}$, define a conditional distribution on the new coordinates by

$$q_C(u \mid z) = \begin{cases} \dfrac{\mu_C(u,z)}{\mu_{S_C}(z)}, & \mu_{S_C}(z) > 0, \\ r_C(u \mid z), & \mu_{S_C}(z) = 0, \end{cases} \qquad u \in \mathcal{A}_{C \setminus S_C},$$

where (u, z) denotes the combined assignment on C and $r_C(\cdot \mid z)$ is any probability distribution. Then

$$w_\mu(a) = \mu_R(a_R) \prod_{C \neq R} q_C(a_{C \setminus S_C} \mid a_{S_C})$$

is a normalized global distribution with bag marginals μ_C. Operationally, draw the root assignment from μ_R and then draw each residual assignment from q_C, conditional on the separator inherited from its parent. This is the sampling construction used by JT-TaS. The proof, including zero separator masses, is given in Lemma 16. A zero-mass separator assignment is reached with probability zero, so the arbitrary conditional on that event cannot change the generated law.

For strictly positive bag marginals, the same distribution has the familiar expression

$$w_\mu(a) = \frac{\displaystyle\prod_{C \in \mathcal{C}_O} \mu_C(a_C)}{\displaystyle\prod_{CD \in \mathcal{E}_O} \mu_{S_{CD}}(a_{S_{CD}})}.$$

The denominator has one factor per tree edge, even when the same separator set occurs on multiple edges. The conditional construction, rather than an undefined ratio, specifies the extension when zeros occur (Lauritzen and Spiegelhalter, 1988b; Wainwright and Jordan, 2008).

What is preserved. The extension w_μ preserves all bag marginals and hence all original factor marginals. It need not reproduce an arbitrary distribution w from which those marginals were obtained, since higher-order dependencies not recorded by the bags need not be identifiable. Nevertheless, this loses no feasible factor allocation in the characteristic-time program, whose dependence on w is entirely through $\tilde{w}$. The JT representation is therefore not a mean-field restriction. For disconnected primal components, empty-separator reconstruction may make the components independent, while preserving every factor marginal.

Why consistency on the original factors is insufficient. Consider three binary agents and pairwise scopes $\{1,2\},\{2,3\},\{1,3\}$. Give each pair probability $1/2$ on $(0,1)$ and $(1,0)$ and zero elsewhere. These distributions agree on every shared singleton, whose marginal is uniform. They cannot, however, be marginals of a global distribution: three binary values cannot be pairwise different. A JT for this triangle has a bag containing all three agents, and its marginal excludes this infeasible collection. The example distinguishes exact JT consistency from an overlap-consistency relaxation on a cyclic family of original factors.

C.7 Computational cost and the choice of elimination order

Graph construction versus table computation. The scopes determine $\mathcal{G}_f$, $\mathcal{G}_p$, the elimination bags, and the JT without evaluating rewards or enumerating joint actions. There are $\sum_e |\mathcal{S}_e|$ factor-graph incidences. Forming the primal graph requires at most $\sum_e \binom{|\mathcal{S}_e|}{2}$ edge-insertion attempts. For a fixed order, writing $d_k = |N_O^+(i_k)|$, fill-in involves at most $\sum_k \binom{d_k}{2}$ pair checks, bounded by $O(Nw_O^2)$. These are graph-operation counts; concrete running times also depend on adjacency maintenance and the procedure used to choose the order.

By contrast, numerical computations enumerate assignments within bags. For dense tables, forming additive bag potentials costs $O(\sum_C |\mathcal{E}_C||\mathcal{A}_C|)$. An upward max-sum pass costs $O(\sum_C(1+|\operatorname{ch}(C)|)|\mathcal{A}_C|)$. Since there are $b_O - 1$ parent-child edges, the total is bounded by

$$O\big((\rho + b_O)K^{w_O+1}\big) \subseteq O\big((\rho + N)K^{w_O+1}\big)\,.$$

The scalar message tables contain at most $\sum_{C\neq R} |\mathcal{A}_{S_C}| \leq (b_O - 1)K^{w_O}$ entries, in addition to stored factor and bag potentials. Keeping all bag tables uses $O(b_O K^{w_O+1})$ entries. These bounds concern tabular arithmetic and storage, not bit complexity. The same separator domains govern sum-product computation (Dechter, 1999; Aji and McEliece, 2000).

The allocation representation has $\sum_C |\mathcal{A}_C| \leq b_O K^{w_O+1}$ nonnegative entries, b_O normalization equalities, and at most $(b_O - 1)K^{w_O}$ scalar separator-consistency equalities, some of which are redundant. Given these tables, root-to-leaf sampling also needs no global-action table: straightforward sampling scans at most $|\mathcal{A}_R| + \sum_{C\neq R} |\mathcal{A}_{C\setminus S_C}| \leq b_O K^{w_O+1}$ conditional entries per draw, apart from coordinate bookkeeping. Precomputing the conditional sampling tables amortizes their construction over repeated draws.

Ordering heuristics. Finding a minimum-width elimination order is NP-hard; the associated treewidth decision problem is NP-complete when the width bound is part of the input (Wainwright and Jordan, 2008). Common heuristics select a variable of minimum current degree or one requiring the fewest fill-in edges, recomputing the criterion in the evolving graph (Koller and Friedman, 2009). These rules need not find a minimum-width order. For example, eliminating the leaves of a star before its center gives width one, whereas eliminating the center first gives width $N - 1$. Thus a poor order can be expensive even when the original graph is a tree. Our bounds retain w_O explicitly and make no optimal-order assumption.

C.8 Relation to the characteristic-time computation

The JT provides two different ingredients in Section 5: exact feasibility through consistent marginals, and exact dynamic programming through separator-conditioned subproblems. Standard max-sum elimination handles an additive reward objective, but does not directly maximize the ratio $c_{\tilde{w}}(a)/\Delta(a)^2$. Our gap-augmented messages retain the cumulative gap in addition to the separator assignment, as specified in Appendix D.3. The extra state is needed for this global numerical dependence; it is not an additional agent variable or a fill-in edge.

Thus w_O controls bag configurations, whereas $H_\theta(O)$ controls cumulative-gap values and can be large even at bounded width. The JT, host assignment, root, and child-processing order are fixed auxiliary choices, suppressed in this notation as in Appendix D.1. The graphical structure can be reused when empirical means or sampling targets change, although numerical potentials and attainable gap states may change. Appendices D.4 and D.5 establish the exact reformulation and its quantization guarantees.

C.9 EXAMPLE: A THREE-AGENT CHAIN

Consider three agents with $\mathcal{A}_i = \{0, 1\}$ and two reward scopes $\mathcal{S}_1 = \{1, 2\}$ and $\mathcal{S}_2 = \{2, 3\}$. The order $O = (3, 2, 1)$ requires no fill-in and yields two maximal bags, $R = \{1, 2\}$ and $C = \{2, 3\}$, joined by separator $S_C = \{2\}$, with $w_O = 1$. Root the JT at R and assign factors 1 and 2 to R and C, respectively.

Optimization. Let $\theta(a) = \theta_1(a_1, a_2) + \theta_2(a_2, a_3)$, with reward tables

$$\theta_1 = \begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix}, \qquad \theta_2 = \begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix},$$

where rows and columns index the first and second arguments, respectively, in the order $0, 1$. Eliminating a_3 gives $m_{C\to R}(a_2) = \max_{a_3} \theta_2(a_2, a_3)$, so $m_{C\to R}(0) = 1$ and $m_{C\to R}(1) = 4$. The root score table is therefore

$$\big[\theta_1(a_1, a_2) + m_{C\to R}(a_2)\big]_{a_1,a_2} = \begin{pmatrix} 4 & 4 \\ 1 & 6 \end{pmatrix}.$$

Its unique maximum selects $(a_1, a_2) = (1, 1)$; backtracking with $a_2 = 1$ selects $a_3 = 1$. Thus $a^\star = (1, 1, 1)$ and $\theta(a^\star) = 6$. In contrast, independently maximizing the two reward factors would select $(0, 0)$ and $(1, 1)$, which disagree on a_2.

Consistent allocation and sampling. For illustration, choose bag marginals

$$\mu_R = \frac{1}{8}\begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}, \qquad \mu_C = \frac{1}{8}\begin{pmatrix} 2 & 2 \\ 1 & 3 \end{pmatrix}.$$

The column sums of μ_R and the row sums of μ_C both give $\mu_{S_C}(0) = \mu_{S_C}(1) = 1/2$, so these marginals are separator-consistent. Their JT extension is

$$w_\mu(a_1, a_2, a_3) = 2\mu_R(a_1, a_2)\mu_C(a_2, a_3).$$

To sample it, first draw $(a_1, a_2) \sim \mu_R$, then draw a_3 with probability $1/2$ of choosing 1 when $a_2 = 0$, and probability $3/4$ when $a_2 = 1$. For example, $w_\mu(1, 1, 1) = (3/8)(3/4) = 9/32$. This construction preserves both prescribed factor marginals without storing a table over all eight joint actions. The chosen marginals illustrate feasibility and sampling; they are not asserted to optimize the characteristic-time program.

D PROOFS FOR JUNCTION-TREE REFORMULATION

We prove Theorem 2, Propositions 3 and 4, Theorem 5, and Proposition 6, in their order of appearance. Section D.1 collects the shared structural and analytic facts. The conic formulation is specified explicitly in Section D.4; its equivalence is proved by equality of projected feasible sets.

Setting and conventions. The action space is the finite Cartesian product $\mathcal{A} = \prod_{i=1}^{N} \mathcal{A}_i$, with $1 \le |\mathcal{A}_i| \le K$ and $|\mathcal{A}| \ge 2$. Factor e has scope $\mathcal{S}_e \subseteq [N]$, local action space $\mathcal{A}_e = \prod_{i \in \mathcal{S}_e} \mathcal{A}_i$, and local mean table $(\theta_{e,x})_{x \in \mathcal{A}_e}$. We use subscripts for table entries, so $\theta_{e,x}$ is the quantity denoted by $\theta_e(x)$ in the original draft. Let $d = \sum_e |\mathcal{A}_e|$ and $\theta(a) = \sum_e \theta_{e,a_e}$. The local means range freely in $\mathbb{R}^d$. The unique maximizer is $a^\star$, and $\Delta(a) = \theta(a^\star) - \theta(a) > 0$ for $a \neq a^\star$. Write $\Delta_{\min} = \min_{a \neq a^\star} \Delta(a)$ and $\Delta_{\max} = \max_{a \neq a^\star} \Delta(a)$. Throughout, $\mu = (\mu_{C,y})$ denotes bag marginals, and $\tilde{w}^{\mu}_{e,x}$ denotes their factor projection. An empty variable set has one assignment; its probability is one. This convention includes empty separators and empty factor scopes.

D.1 AUXILIARY LEMMAS

Lemma 13 (Elimination bags and a compatible junction tree)**.** *Let $O = (v_1, \dots, v_N)$ be any elimination order of the primal graph. Let S_i be the remaining neighbors of v_i immediately before its elimination, including fill edges introduced earlier, and set $C_i = \{v_i\} \cup S_i$. Then $w_O = \max_i |S_i|$. If $S_i \neq \varnothing$ and $p(i) = \min\{j : v_j \in S_i\}$, then*

$$p(i) > i, \qquad S_i \subseteq C_{p(i)}, \qquad C_i \cap C_{p(i)} = S_i. \tag{15}$$

The edges $C_i C_{p(i)}$ form a running-intersection forest. Joining its components by empty-separator edges gives a junction tree. Removing nonmaximal bags by edge contractions gives a junction tree on the inclusion-maximal elimination bags, with at most N bags, each of size at most $w_O + 1$, and separators of size at most w_O. Every factor scope is contained in a bag.

Proof. This is the elimination construction underlying exact graphical inference (Dechter, 1999; Wainwright and Jordan, 2008). We give the details for arbitrary O, including disconnected primal graphs.

Eliminating v_i connects every pair in S_i. Fix $v_j \in S_i \setminus \{v_{p(i)}\}$. By the minimality of $p(i)$, $j > p(i)$. The edge $v_j v_{p(i)}$ is present after eliminating v_i and remains present until $v_{p(i)}$ is eliminated. Hence $v_j \in S_{p(i)}$. This proves $S_i \subseteq C_{p(i)}$. Since v_i does not belong to any later elimination bag, the intersection in (15) is exactly S_i. Parent indices strictly increase, so the parent graph is acyclic.

Fix a variable v_j. If $v_j \in C_i$ with $i < j$, then $v_j \in S_i$. Equation (15) implies $v_j \in C_{p(i)}$ and $i < p(i) \le j$. If $p(i) < j$, repeat. The strictly increasing parent chain reaches C_j, and every bag on that chain contains v_j. Thus

$$v_j \in C_i \quad \Longrightarrow \quad \text{the path from } C_i \text{ to } C_j \text{ contains } v_j \text{ in every bag.} \tag{16}$$

This proves running intersection. It also shows that the variable sets of distinct forest components are disjoint: a shared variable would connect both components to its own elimination bag. They may therefore be connected by arbitrary tree edges with empty separators.

Suppose a bag B is contained in another bag D. Running intersection implies that every bag on their path contains B. In particular, the first neighbor B' toward D contains B. Contract the edge BB' by deleting B and attaching its other neighbors to B'. For any variable occurring on a path through B, that variable also occurs in B'; thus the contraction preserves running intersection. Coverage is preserved because $B \subseteq B'$. Repeating this operation leaves precisely the inclusion-maximal bags. No bag is enlarged, and no bag is created. Consequently there are at most N remaining bags and their sizes are at most $w_O + 1$. Adjacent distinct maximal bags cannot contain each other, so their intersection has size at most w_O.

For a nonempty factor scope, let v_i be its earliest eliminated variable. The scope is a clique in the primal graph, hence $\mathcal{S}_e \setminus \{v_i\} \subseteq S_i$ and $\mathcal{S}_e \subseteq C_i$. Subsequent contractions retain a bag containing it. An empty scope is contained in every bag. This proves the final assertion. □

Fix henceforth the induced junction tree $\mathcal{T}_O = (\mathcal{C}_O, \mathcal{E}_O)$, a root R, a host bag $C(e) \supseteq \mathcal{S}_e$ for each factor, and a child order at each bag. Write $b_O = |\mathcal{C}_O| \le N$. For a bag C, let $P(C)$ be its parent,

$S_C = C \cap P(C)$, and U_C the union of the bags in its descendant subtree, including C. Set $S_R = \varnothing$. These auxiliary choices are fixed; the notation $H_\theta(O)$ suppresses them.

Lemma 14 (Subtree separation and gluing)**.** *For every nonroot bag C, the variables shared by U_C and bags outside its subtree are exactly S_C. If D, D' are distinct children of C, then $(U_D \setminus C) \cap (U_{D'} \setminus C) = \varnothing$. Consequently, fixing $y \in \mathcal{A}_C$, any child-subtree assignments that agree with y on their separators combine into a unique assignment on U_C.*

Proof. If a variable occurs in a descendant bag of C and a bag outside that subtree, their path crosses the edge $CP(C)$. Running intersection forces the variable into both endpoint bags, hence into S_C. The reverse inclusion holds because S_C belongs to both C and $P(C)$. Likewise, a variable shared by two different child subtrees must occur on their connecting path and therefore in C. After y is fixed, no residual variable of one child subtree can conflict with a residual variable of another. Each child agrees with y on its entire intersection with C, so their union is well defined and unique. □

Lemma 15 (Finite objective and attainment)**.** *The set $\tilde{\Lambda}$ is nonempty and compact. The extended-valued functional T_θ is convex and lower semicontinuous on it. It is finite exactly when $\tilde{w}_{e,x} > 0$ for every factor with $|\mathcal{A}_e| \geq 2$ and every $x \in \mathcal{A}_e$. Its minimum is attained and satisfies*

$$0 < \frac{2}{\Delta_{\max}^2} \leq T_\theta^\star \leq \frac{2d}{\Delta_{\min}^2} < \infty. \tag{17}$$

Proof. By definition, $\tilde{\Lambda}$ is a linear image of the finite-dimensional probability simplex Λ; it is therefore compact and nonempty. The extended reciprocal $z \mapsto 1/z$ is convex and lower semicontinuous on $[0, \infty)$. Positive finite sums and a finite maximum preserve these properties, which proves the assertion for T_θ.

For a nontrivial factor and $x \neq a_e^\star$, every joint extension of x is a competitor whose numerator contains $\tilde{w}_{e,x}^{-1}$. For $x = a_e^\star$, extend any other local action of that factor; its numerator contains $\tilde{w}_{e,a_e^\star}^{-1}$. Thus a zero marginal in any of these coordinates forces $T_\theta(\tilde{w}) = \infty$. Conversely, positivity of these coordinates makes every competitor numerator finite. Singleton factor marginals equal one and occur in no competitor numerator.

Independent uniform agent actions induce $\tilde{w}_{e,x}^0 = 1/|\mathcal{A}_e|$, hence $D_{\tilde{w}^0}(a) \leq 2 \sum_e |\mathcal{A}_e| = 2d$. This gives the upper bound in (17). Lower semicontinuity and compactness give attainment. Every competitor differs from $a^\star$ on at least one factor, since otherwise their total rewards would agree. Each probability is at most one, so its numerator is at least two. This proves the positive lower bound. □

D.2 Proof of Proposition 3: globally feasible allocations

We use the local-to-global consistency theorem for junction trees (Wainwright and Jordan, 2008, Section 2.5, Proposition 1 and equations (2.10)–(2.11)). The following construction proves the needed specialization without dividing by a zero separator probability.

Lemma 16 (Host-independent projections and zero-safe reconstruction)**.** *For $\mu \in \mathcal{M}_O$, the factor projection $\tilde{w}_{e,x}^\mu = \sum_{y \in \mathcal{A}_{C(e)} : y_{\mathcal{S}_e} = x} \mu_{C(e),y}$ is independent of the selected host. There exists a probability distribution w_μ on $\mathcal{A}$ whose bag marginals are exactly μ.*

Proof. Host independence. If two bags contain $\mathcal{S}_e$, then every bag and every separator on their path also contains $\mathcal{S}_e$, by running intersection. For adjacent bags C, D on that path,

$$\mu_C^{\mathcal{S}_e} = (\mu_C^{C \cap D})^{\mathcal{S}_e} = (\mu_D^{C \cap D})^{\mathcal{S}_e} = \mu_D^{\mathcal{S}_e}. \tag{18}$$

Iteration proves equality at the two endpoints.

Conditional kernels. For a child C with separator S_C, let $\mu_{S_C,z}$ be its common separator mass. Choose an arbitrary probability vector $r_{C,z}$ on $\mathcal{A}_{C \setminus S_C}$ and define

$$q_C(y_{C \setminus S_C} \mid z) = \begin{cases} \mu_{C,y} / \mu_{S_C,z}, & \mu_{S_C,z} > 0,\ y_{S_C} = z, \\ r_{C,z}(y_{C \setminus S_C}), & \mu_{S_C,z} = 0. \end{cases} \tag{19}$$

Both branches sum to one. In the zero-mass branch, nonnegativity gives $\mu_{C,y} = 0$ for every extension y of z.

Draw the root assignment from μ_R and process its descendants in parent-before-child order, using (19). Lemma 14 ensures that every new coordinate is assigned only once and all shared coordinates have already been fixed by the parent separator. Consequently this procedure defines the normalized law

$$w_\mu(a) = \mu_{R,a_R} \prod_{C \neq R} q_C(a_{C \setminus S_C} \mid a_{S_C}). \tag{20}$$

Normalization also follows by successively summing the normalized leaf kernels.

Bag marginals. The root has its prescribed marginal. Suppose the parent of C has already been generated with its prescribed marginal. Separator consistency makes the probability of its separator assignment z equal to $\mu_{S_C,z}$. Thus for $y \in \mathcal{A}_C$,

$$\mathbb{P}_{w_\mu}(A_C = y) = \begin{cases} \mu_{S_C,y_{S_C}}\, \mu_{C,y}/\mu_{S_C,y_{S_C}} = \mu_{C,y}, & \mu_{S_C,y_{S_C}} > 0, \\ 0 = \mu_{C,y}, & \mu_{S_C,y_{S_C}} = 0. \end{cases} \tag{21}$$

Adding descendants cannot change previously generated marginals because each new kernel is normalized. Induction proves the assertion for every bag, including across empty separators. □

Proof of Proposition 3. Let $\tilde{w} \in \tilde{\Lambda}$ and choose $w \in \Lambda$ inducing it. Its bag marginals $\mu_{C,y} = \sum_{a:a_C=y} w_a$ are nonnegative, normalized, and satisfy $\mu_C^{S_{CD}}(z) = \sum_{a:a_{S_{CD}}=z} w_a = \mu_D^{S_{CD}}(z)$. Hence $\mu \in \mathcal{M}_O$, and

$$\tilde{w}^\mu_{e,x} = \sum_{y:y_{\mathcal{S}_e}=x} \sum_{a:a_{C(e)}=y} w_a = \sum_{a:a_e=x} w_a = \tilde{w}_{e,x}. \tag{22}$$

This proves $\tilde{\Lambda} \subseteq \{\tilde{w}^\mu : \mu \in \mathcal{M}_O\}$. Conversely, Lemma 16 extends each $\mu \in \mathcal{M}_O$ to a global w_μ. Its factor marginals are $\tilde{w}^\mu$, which therefore belong to $\tilde{\Lambda}$. Both inclusions prove equality, with neither a restriction nor a relaxation.

There are $\sum_C |\mathcal{A}_C| \le b_O K^{w_O+1}$ bag entries. Their description needs b_O normalization equalities, $\sum_{CD \in \mathcal{E}_Q} |\mathcal{A}_{S_{CD}}| \le (b_O - 1)K^{w_O}$ separator equalities, and nonnegativity. These counts will be used in the conic formulation below. □

D.3 Proof of Proposition 4: the competitor maximization

Fix $\tilde{w} \in \tilde{\Lambda}$. Let $\Delta_{e,x} = \theta_{e,a_e^\star} - \theta_{e,x}$, so $\Delta(a) = \sum_e \Delta_{e,a_e}$ and $D_{\tilde{w}}(a) = \sum_e c_{e,\tilde{w},a_e}$. The competitor ratio is not an additive bag objective: its denominator is the square of the sum of all local gaps. Conditioning on that sum yields the exact reduction

$$\begin{aligned} \Phi_{\theta,\tilde{w}}(s) &= \max_{a \in \mathcal{A}:\Delta(a)=s} D_{\tilde{w}}(a), \\ T_\theta(\tilde{w}) &= \max_{s \in \mathcal{D}_R(\varnothing),\ s>0} \frac{\Phi_{\theta,\tilde{w}}(s)}{s^2}, \qquad \mathcal{D}_R(\varnothing) = \{\Delta(a) : a \in \mathcal{A}\}. \end{aligned} \tag{23}$$

The construction below augments JT elimination with this additive statistic. Constrained message passing with global sufficient statistics is developed by Bauer et al. (2023, Section 4, Algorithm 1 and Theorem 1); here we give the exact recurrence and sequential-merge count needed for the characteristic-time program. No complexity assertion from that general theorem is substituted for the count below.

Bag contributions and reachable states. Let $\mathcal{F}_C = \{e : C(e) = C\}$ and let $\mathcal{F}_{\downarrow C}$ be the factors hosted in the subtree of C. Define $g_{C,y} = \sum_{e \in \mathcal{F}_C} \Delta_{e,y_{\mathcal{S}_e}}$ and $c_{C,y} = \sum_{e \in \mathcal{F}_C} c_{e,\tilde{w},y_{\mathcal{S}_e}}$. Every factor is counted once. For $z \in \mathcal{A}_{S_C}$, set

$$\begin{aligned} \mathcal{D}_C(z) &= \left\{ \sum_{e \in \mathcal{F}_{\downarrow C}} \Delta_{e,a_e} : a \in \mathcal{A}_{U_C},\ a_{S_C} = z \right\}, \\ M_{C,z,s} &= \max_{\substack{a \in \mathcal{A}_{U_C},\ a_{S_C}=z \\ \sum_{e \in \mathcal{F}_{\downarrow C}} \Delta_{e,a_e} = s}} \sum_{e \in \mathcal{F}_{\downarrow C}} c_{e,\tilde{w},a_e}, \qquad s \in \mathcal{D}_C(z). \end{aligned} \tag{24}$$

Only reachable entries are stored. The set $\mathcal{D}_C(z)$ is finite and nonempty for every separator assignment because local action spaces are Cartesian. It is defined independently of $\tilde{w}$; a reachable message may equal $+\infty$ when an allocation has zero relevant marginals. An unreachable state is excluded rather than added to a possibly infinite cost.

Let $D_{C,1}, \dots, D_{C,m_C}$ be the ordered children of C. With $D = D_{C,j}$ and $S_D = C \cap D$, define

$$\begin{aligned} \Gamma_{C,0}(y) &= \{g_{C,y}\}, \\ \Gamma_{C,j}(y) &= \{p + q : p \in \Gamma_{C,j-1}(y),\ q \in \mathcal{D}_D(y_{S_D})\}, \qquad j \in [m_C], \\ \mathcal{D}_C(z) &= \bigcup_{y \in \mathcal{A}_C :\, y_{S_C} = z} \Gamma_{C,m_C}(y). \end{aligned} \tag{25}$$

The corresponding numerical recursion is

$$\begin{aligned} A_{C,0,y,g_{C,y}} &= c_{C,y}, \\ A_{C,j,y,s} &= \max_{\substack{p \in \Gamma_{C,j-1}(y),\ q \in \mathcal{D}_D(y_{S_D}) \\ p+q=s}} \{A_{C,j-1,y,p} + M_{D,y_{S_D},q}\}, \\ M_{C,z,s} &= \max_{\substack{y \in \mathcal{A}_C :\, y_{S_C} = z \\ s \in \Gamma_{C,m_C}(y)}} A_{C,m_C,y,s}. \end{aligned} \tag{26}$$

For a leaf, $m_C = 0$ and the last line acts directly on the base entries. For the root, $z = \varnothing$ is the sole separator assignment. Signed intermediate gaps, including negative and zero values, are retained throughout.

Lemma 17 (Correctness of the sequential recurrence)**.** *Equations* (25)–(26) *compute precisely the supports and values in* (24)*. At the root, $M_{R,\varnothing,s} = \Phi_{\theta,\tilde{w}}(s)$ for every attainable s, including $s = 0$.*

Proof. For $j \in \{0, \dots, m_C\}$, let $\mathcal{F}_{C,j} = \mathcal{F}_C \cup \bigcup_{\ell \le j} \mathcal{F}_{\downarrow D_{C,\ell}}$ and $V_{C,j} = C \cup \bigcup_{\ell \le j} U_{D_{C,\ell}}$. We prove that $\Gamma_{C,j}(y)$ is the set of sums $\sum_{e \in \mathcal{F}_{C,j}} \Delta_{e,a_e}$ over assignments to $V_{C,j}$ restricting to y on C, and $A_{C,j,y,s}$ is the maximum corresponding cost at fixed sum s.

For $j = 0$, the hosted factors are fixed by y, so the only sum is $g_{C,y}$ and its cost is $c_{C,y}$. This proves the base case. Assume the claim for the first $j - 1$ children and assume the child message has the semantics in (24). Every assignment for stage j restricts to an assignment for stage $j - 1$ and one for the subtree of $D = D_{C,j}$. Their gaps add, their hosted factor sets are disjoint, and their costs add. Thus its gap belongs to the sumset in (25), and its cost does not exceed the associated candidate in (26).

Conversely, fix a pair (p, q) considered by that recurrence. Choose maximizing completions for the preceding stage and the child message. These maxima exist because each feasible assignment set is finite, also when its value is $+\infty$. By Lemma 14, their residual variables are disjoint outside C, and both agree with y on their intersections with C. They therefore glue into an assignment at stage j with exactly the sum $p + q$ and the sum of the two costs. This proves the reverse inequality and proves that every element of the computed support is genuinely attainable.

Finally, maximizing over the full bag assignment y at fixed z eliminates the variables in $C \setminus S_C$, giving exactly (24). Induction over child stages and then from leaves to root proves both recurrences. At the root, $\mathcal{F}_{\downarrow R} = [\rho]$ and $U_R = [N]$, so the semantic maximum is exactly the gap profile. □

Definition of the width. For this fixed elimination scheme, define

$$H_\theta(O) = \max\left\{1,\ \max_{C \in \mathcal{C}_O,\ z \in \mathcal{A}_{S_C}} |\mathcal{D}_C(z)|,\ \max_{\substack{C \in \mathcal{C}_O,\ 0 \le j \le m_C \\ y \in \mathcal{A}_C}} |\Gamma_{C,j}(y)|\right\}. \tag{27}$$

Thus the width bounds every intermediate combination table, not only the final separator messages. It is finite for arbitrary real local means and depends on θ and the fixed decomposition choices, not on $\tilde{w}$.

Lemma 18 (Arithmetic and storage of the gap computation)**.** *Given the bag-local tables $(g_{C,y}, c_{C,y})$, the profile can be computed with $O(NK^{w_O+1}H_\theta(O)^2)$ arithmetic operations and exact comparisons, using $O(NK^{w_O+1}H_\theta(O))$ numerical table storage. Constructing the bag-local tables from factor inputs costs additionally $O(\rho K^{w_O+1})$ operations and at most that much input storage.*

Proof. Write $H = H_\theta(O)$ and $L = K^{w_O+1}$. There are at most $b_O L$ bag assignments. Since $\sum_C m_C = b_O - 1$,

$$\sum_C (m_C + 1) = 2b_O - 1, \qquad \sum_{C,j,y} |\Gamma_{C,j}(y)| \le (2b_O - 1)LH, \qquad \sum_{C,z} |\mathcal{D}_C(z)| \le b_O LH. \tag{28}$$

Each merge for a fixed y considers at most H^2 pairs (p, q). There are $(b_O - 1)L$ such bag-assignment/child incidences at most. Projection to separator messages costs $O(b_O LH)$, and base cases cost $O(b_O L)$. This gives $O(b_O LH^2)$ operations once identical gap labels are combined.

This count can include exact support construction without assuming constant-time hashing of real numbers. Store supports in increasing order. For each fixed p, the list $p + \mathcal{D}_D(y_{S_D})$ is ordered and has at most H entries. Merge it with the accumulated output list, retaining the larger value on equal keys. The accumulated list is a subset of the final output support and also has at most H entries. Each merge therefore uses $O(H)$ additions/comparisons, and at most H shifted lists give $O(H^2)$ work and $O(H)$ temporary storage. The separator projection similarly merges sorted stage lists, each with an accumulated support of size at most H. This implements the count deterministically in exact arithmetic, with no hidden factor exponential in m_C.

Equation (28) bounds storage for all stage and separator tables. A full list of the $O(b_O LH^2)$ transitions need not be retained during evaluation. Constructing each bag-local contribution sums its hosted factors for each bag assignment, costing at most $\sum_e |\mathcal{A}_{C(e)}| \le \rho L$ operations. Substituting $b_O \le N$ proves the claim. Bit costs of representing and comparing arbitrary real labels are not included in an arithmetic-operation bound. □

Proof of Proposition 4. Lemma 17 computes the entire profile exactly. Because $a^\star$ is unique, the root gap is zero only at $a^\star$ and is positive at every competitor. Grouping the finite competitor set by its total gap gives (23). Lemma 18 gives the claimed time and memory after bag-local contributions are available; including their construction gives, respectively, $O((NH_\theta(O)^2 + \rho)K^{w_O+1})$ and $O((NH_\theta(O) + \rho)K^{w_O+1})$. The support recursion contains only gaps and scopes, so the reachable supports and transition incidences do not change when $\tilde{w}$ changes. □

D.4 Proof of Theorem 5: equivalent SOCP and size

All reachable supports below are those of (25) for the fixed instance θ. They are constants of the optimization, including their numerical gap labels. We introduce variables $\mu_{C,y} \ge 0$, factor marginals $p_{e,x} \ge 0$, reciprocal bounds $u_{e,x} \ge 0$, a scalar $t \ge 0$, stage variables $V_{C,j,y,s} \in \mathbb{R}$, and message variables $U_{C,z,s} \in \mathbb{R}$. The stage and message variables exist only at reachable states. Here p is an auxiliary copy of $\tilde{w}^\mu$, not an additional sampling family.

Define the rotated cone $\mathcal{Q}_r^3 = \{(r, s, v) : r, s \ge 0,\ 2rs \ge v^2\}$. Its use for reciprocal epigraphs is elementary conic modeling (Boyd and Vandenberghe, 2004): $(p, u, \sqrt{2}) \in \mathcal{Q}_r^3$ is equivalent to $p, u \ge 0$ and $pu \ge 1$. Equivalently, the ordinary SOC inequality is $\|(2, p - u)\|_2 \le p + u$. Both equivalences follow by squaring and simplifying.

Let $c_{e,u,x} = 0$ for $x = a_e^\star$, and $c_{e,u,x} = u_{e,a_e^\star} + u_{e,x}$ otherwise. Put $c_{C,y}(u) = \sum_{e \in \mathcal{F}_C} c_{e,u,y_{\mathcal{S}_e}}$, which is affine in u. The complete extended formulation minimizes t subject to the following constraints:

$$\begin{aligned}
\sum_{y \in \mathcal{A}_C} \mu_{C,y} &= 1 && (C \in \mathcal{C}_O),\\
\sum_{y : y_{S_{CD}} = z} \mu_{C,y} &= \sum_{y' : y'_{S_{CD}} = z} \mu_{D,y'} && (CD \in \mathcal{E}_O,\ z \in \mathcal{A}_{S_{CD}}),\\
p_{e,x} &= \sum_{y : y_{\mathcal{S}_e} = x} \mu_{C(e),y} && (e \in [\rho],\ x \in \mathcal{A}_e),\\
(p_{e,x}, u_{e,x}, \sqrt{2}) &\in \mathcal{Q}_r^3 && (e \in [\rho],\ x \in \mathcal{A}_e),\\
V_{C,0,y,g_{C,y}} &= c_{C,y}(u) && (C \in \mathcal{C}_O,\ y \in \mathcal{A}_C).
\end{aligned} \tag{29}$$

For every C, $j \in [m_C]$, $D = D_{C,j}$, $y \in \mathcal{A}_C$, $p' \in \Gamma_{C,j-1}(y)$, and $q \in \mathcal{D}_D(y_{S_D})$, impose

$$V_{C,j,y,p'+q} \ge V_{C,j-1,y,p'} + U_{D,y_{S_D},q}. \tag{30}$$

For every C, $y \in \mathcal{A}_C$, and $s \in \Gamma_{C,m_C}(y)$, impose

$$U_{C,y_{S_C},s} \geq V_{C,m_C,y,s}, \qquad U_{R,\varnothing,s} \leq ts^2 \quad (s \in \mathcal{D}_R(\varnothing),\ s > 0). \tag{31}$$

The root bound in (31) is a separate constraint for each positive root state. No bound of this type is imposed on signed intermediate states. All displayed equalities and inequalities except the constant-dimensional rotated cones are affine in the optimization variables: in particular s^2 is a fixed coefficient, not a variable.

Lemma 19 (Epigraph lifting of the gap computation)**.** *Fix finite reciprocal variables u. Let A^u, M^u be the numerical recursion* (26) *with local costs $c_{C,y}(u)$. Every feasible collection (V, U) satisfying the base, merge, and projection constraints obeys $V \geq A^u$ and $U \geq M^u$ coordinatewise. Conversely, choosing $V = A^u$ and $U = M^u$ satisfies all these constraints. There exists a collection satisfying these constraints and the root bounds if and only if $M^u_{R,\varnothing,s} \leq ts^2$ for every positive root state.*

Proof. At every base entry, $V_{C,0,y,g_{C,y}} = A^u_{C,0,y,g_{C,y}}$ by definition. Process the recursion from leaves to root and child stages in increasing order. Assume the preceding stage and the relevant child message satisfy their coordinatewise lower bounds. For every compatible (p', q),

$$V_{C,j,y,p'+q} \geq V_{C,j-1,y,p'} + U_{D,y_{S_D},q} \geq A^u_{C,j-1,y,p'} + M^u_{D,y_{S_D},q}. \tag{32}$$

Maximizing over the compatible pairs yields $V_{C,j,y,s} \geq A^u_{C,j,y,s}$. The projection inequalities similarly imply $U_{C,z,s} \geq \max_{y:y_{S_C}=z,\ s\in\Gamma_{C,m_C}(y)} A^u_{C,m_C,y,s} = M^u_{C,z,s}$. The acyclic dependency order makes this induction exhaustive.

Conversely, the exact maximum is at least each of its candidate values. Thus setting every variable equal to its numerical recurrence value satisfies all base equalities and all merge and projection inequalities simultaneously. The root upper bounds are therefore feasible precisely when they bound these least feasible root values. This is a feasibility statement, not an assertion that every epigraph variable must be tight at an optimum. □

Lemma 20 (Exact equality of projected feasible sets)**.** *Let $\mathcal{K}_\theta$ be the feasible set defined by* (29)–(31) *and nonnegativity of μ, p, u, t. Then*

$$\mathrm{proj}_{(p,t)}\, \mathcal{K}_\theta = \{(\tilde{w}, t) : \tilde{w} \in \tilde{\Lambda},\ t \in \mathbb{R}_+,\ T_\theta(\tilde{w}) \leq t\}. \tag{33}$$

Proof. Extended feasibility implies original feasibility. Take $(\mu, p, u, V, U, t) \in \mathcal{K}_\theta$. The first three groups of constraints and Proposition 3 imply $p = \tilde{w}^\mu \in \tilde{\Lambda}$. The cones imply $p_{e,x} > 0$ and $u_{e,x} \geq 1/p_{e,x}$. Hence $c_{e,u,x} \geq c_{e,p,x}$ and, for every global action, $\sum_e c_{e,u,a_e} \geq D_p(a)$. For every positive root state,

$$\Phi_{\theta,p}(s) \leq M^u_{R,\varnothing,s} \leq U_{R,\varnothing,s} \leq ts^2. \tag{34}$$

The middle inequality is Lemma 19. Equation (23) therefore gives $T_\theta(p) \leq t$.

Original feasibility admits an exact lift. Conversely, take $\tilde{w} \in \tilde{\Lambda}$ with $T_\theta(\tilde{w}) \leq t < \infty$. Proposition 3 gives a $\mu \in \mathcal{M}_O$ with $\bar{w}^\mu = \tilde{w}$. Lemma 15 makes all nontrivial factor marginals strictly positive; singleton marginals are one. Set $p = \tilde{w}$, $u_{e,x} = 1/p_{e,x}$, and let V, U equal the exact recurrence for these costs. The cones hold with equality, and Lemma 19 gives every recursion constraint. At the root, $U_{R,\varnothing,s} = \Phi_{\theta,\tilde{w}}(s) \leq ts^2$, so the lift is feasible. This proves both inclusions in (33), including the absence of finite lifts at allocations with infinite objective. □

Proof of Theorem 5. All reachable supports are finite, even for arbitrary real local means. The displayed extended formulation is therefore a finite SOCP. Lemma 20 proves equality of its optimal value with $T^\star_\theta$. Lemma 15 provides an attained finite optimizer of the original problem; the reverse lifting construction attains the same value in the SOCP. Every optimal SOCP solution projects to $p \in \tilde{\Lambda}$ with $T_\theta(p) \leq t = T^\star_\theta$, and hence supplies an information-optimal factor allocation.

It remains to prove the size, including the intermediate variables that make the quadratic dependence on H valid at high-degree bags. Let $H = H_\theta(O)$ and $L = K^{w_O+1}$. The bag variables number at most $b_O L$, the p, u variables number $2d \leq 2\rho L$, the scalar t contributes one, and (28) gives

$$\begin{aligned} \#V &\leq (2b_O - 1)LH, & \#U &\leq b_O LH, \\ \#\text{merge inequalities} &\leq (b_O - 1)LH^2, & \#\text{projection inequalities} &\leq b_O LH. \end{aligned} \tag{35}$$

There are at most $b_O L$ base equalities, b_O normalizations, $(b_O - 1)K^{w_O}$ separator equalities, d factor-projection equalities, d constant-dimensional cones, and H root upper bounds. Nonnegativity adds $O(b_O L + \rho L)$ scalar inequalities. Since $H \geq 1$ and $b_O \leq N$, the totals are

$$\#\text{scalar variables} = O\big((NH + \rho)K^{w_O+1}\big), \qquad \#\text{constraints} = O\big((NH^2 + \rho)K^{w_O+1}\big). \quad (36)$$

These counts do not hide dense global-action rows. Each merge row has three variable occurrences. Across all factor-projection rows, each host-bag assignment contributes once for each hosted factor, giving $\sum_e |\mathcal{A}_{C(e)}| \leq \rho L$ occurrences. Across all separator rows, the number of occurrences is $\sum_{CD}(|\mathcal{A}_C| + |\mathcal{A}_D|) \leq 2(b_O - 1)L$. The base-cost rows similarly have $O(\rho L)$ occurrences. Thus the sparse coefficient representation has the same stated order, up to constant-dimensional cone blocks.

The theorem is an exact equivalence and formulation-size statement. It does not assert that H is polynomial in N, that arbitrary real gap labels have a finite bit representation, or that a numerical solver returns an exact symbolic optimizer. No lattice assumption is needed for the equivalence itself. □

D.5 Proof of Proposition 6: certified approximation

Lemma 21 (Lattice width for all reachable tables)**.** *Suppose $\theta_{e,x} \in \gamma\mathbb{Z}$ for every local entry, where $\gamma > 0$, and let $B_e = \max_x \theta_{e,x} - \min_x \theta_{e,x}$. Then the width in* (27) *satisfies*

$$H_\theta(O) \leq 1 + \left\lfloor \frac{\sum_e B_e}{\gamma} \right\rfloor \leq 1 + \gamma^{-1} \sum_e B_e. \quad (37)$$

Proof. Every stage and separator support corresponds to a fixed subset F of hosted factors: respectively $F = \mathcal{F}_{C,j}$ or $F = \mathcal{F}_{\downarrow C}$. Its gap is a sum of the form $\sum_{e \in F}(\theta_{e,a_e^\star} - \theta_{e,a_e})$. For each factor, this contribution is in $\gamma\mathbb{Z}$ and lies in $[\theta_{e,a_e^\star} - \max_x \theta_{e,x},\ \theta_{e,a_e^\star} - \min_x \theta_{e,x}]$, an interval of length B_e. Therefore

$$\sum_{e \in F}(\theta_{e,a_e^\star} - \theta_{e,a_e}) \in \gamma\mathbb{Z} \cap \left[\sum_{e \in F}(\theta_{e,a_e^\star} - \max_x \theta_{e,x}),\ \sum_{e \in F}(\theta_{e,a_e^\star} - \min_x \theta_{e,x}) \right]. \quad (38)$$

The interval has length $\sum_{e \in F} B_e \leq \sum_e B_e$. It contains at most $1 + \lfloor \sum_e B_e / \gamma \rfloor$ lattice points. Fixing bag or separator assignments can only remove attainable values. Taking the maximum over all stages and messages proves (37). In particular, the argument does not replace signed gaps by nonnegative ones, and it does not require a factorwise optimal $a_e^\star$. □

Lemma 22 (Uniform effect of nearest-grid rounding)**.** *Let $\theta^\eta_{e,x} = \eta\,\mathrm{round}(\theta_{e,x}/\eta)$ and $q = \rho\eta/\Delta_{\min} < 1$. Then $a^\star$ remains the unique optimum, and for all $\tilde{w} \in \tilde{\Lambda}$,*

$$\frac{T_\theta(\tilde{w})}{(1+q)^2} \leq T_{\theta^\eta}(\tilde{w}) \leq \frac{T_\theta(\tilde{w})}{(1-q)^2}, \qquad \frac{T^\star_\theta}{(1+q)^2} \leq T^\star_{\theta^\eta} \leq \frac{T^\star_\theta}{(1-q)^2}. \quad (39)$$

The pointwise inequalities use the extended-value convention when the objective is infinite.

Proof. Let $r_{e,x} = \theta^\eta_{e,x} - \theta_{e,x}$, so $|r_{e,x}| \leq \eta/2$. For any action a, $|\theta^\eta(a) - \theta(a)| \leq \rho\eta/2$. More precisely, cancellation on unchanged factors gives

$$|\theta^\eta(a^\star) - \theta^\eta(a) - \Delta(a)| = \left| \sum_{e : a_e \neq a_e^\star} (r_{e,a_e^\star} - r_{e,a_e}) \right| \leq \rho\eta = q\Delta_{\min} \leq q\Delta(a). \quad (40)$$

Thus $\theta^\eta(a^\star) - \theta^\eta(a) \geq (1-q)\Delta(a) > 0$. The optimum is preserved and $(1-q)\Delta(a) \leq \Delta^\eta(a) \leq (1+q)\Delta(a)$. The factor observations and the best action have not changed, so the numerator $D_{\tilde{w}}(a)$ is identical in the two characteristic-time objectives. Divide it by the squared gap inequalities and maximize over all competitors to obtain the pointwise bounds. If a numerator is infinite, the same zero marginal makes both objectives infinite. Infimizing the pointwise inequalities over the unchanged feasible set proves the optimal-value bounds. □

Proof of Proposition 6. Let $\varepsilon \in (0, 1]$ and choose precisely the resolution in the main text, $\eta_\varepsilon = \varepsilon\Delta_{\min}/(8\rho)$. For this proof alone write $q = \rho\eta_\varepsilon/\Delta_{\min} = \varepsilon/8 \le 1/8$. Lemma 22 preserves the unique best action and gives

$$T^\star_\theta \le (1+q)^2 T^\star_{\theta^\varepsilon} = \overline{T}_{\theta,\varepsilon} \le \left(\frac{1+q}{1-q}\right)^2 T^\star_\theta. \tag{41}$$

The last factor is bounded without an implicit choice of constants:

$$\left(\frac{1+\varepsilon/8}{1-\varepsilon/8}\right)^2 = 1 + \frac{\varepsilon}{2(1-\varepsilon/8)^2} \le 1 + \frac{32}{49}\varepsilon \le 1+\varepsilon. \tag{42}$$

This proves the claimed one-sided certificate $T^\star_\theta \le \overline{T}_{\theta,\varepsilon} \le (1+\varepsilon)T^\star_\theta$, not merely a two-sided approximation of the uncorrected quantized value.

For the state bound, put $B^\varepsilon_e = \max_x \theta^\varepsilon_{e,x} - \min_x \theta^\varepsilon_{e,x}$. Lemma 21, applied at spacing η_ε, gives exactly the stated inequality. The original factor ranges also give an explicit bound independent of the new range notation:

$$\begin{aligned} H_{\theta^\varepsilon}(O) &\le 1 + \frac{8\rho}{\varepsilon\Delta_{\min}} \sum_e B^\varepsilon_e, \\ B^\varepsilon_e \le B_e + \eta_\varepsilon \quad &\Longrightarrow \quad H_{\theta^\varepsilon}(O) \le 1 + \rho + \frac{8\rho}{\varepsilon\Delta_{\min}} \sum_e B_e. \end{aligned} \tag{43}$$

The range inequality follows because rounding each endpoint changes it by at most $\eta_\varepsilon/2$. Thus, for fixed instance parameters and decomposition, $H_{\theta^\varepsilon}(O) = O(\varepsilon^{-1})$. Substitution into (36) gives the stated $O(\varepsilon^{-1})$ and $O(\varepsilon^{-2})$ dependence of the numbers of variables and constraints.

The same proof also certifies the allocation returned by the quantized solve. If $\tilde{w}^\varepsilon$ minimizes T_{θ^ε}, the pointwise bound in (39) yields $T^\star_\theta \le T_\theta(\tilde{w}^\varepsilon) \le (1+q)^2 T^\star_{\theta^\varepsilon} = \overline{T}_{\theta,\varepsilon} \le (1+\varepsilon)T^\star_\theta$. No new approximation of the feasible allocation set is used. □

Interpretation of the state bound. For $B_e \le B$, Lemma 21 gives $H_\theta(O) \le 1 + \rho B/\gamma$; integer-valued means correspond to $\gamma = 1$. At fixed induced width the resulting formulation size is pseudo-polynomial in the numerical range measured in lattice units, not necessarily polynomial in its binary encoding. Bounded graphical width alone does not bound H: with N binary agents, unary factors, $\theta_{i,0} = 0$, and $\theta_{i,1} = -2^{i-1}$, the unique optimum is 0^N, the induced width is zero, and the root gap set is $\{0, 1, \dots, 2^N - 1\}$. Hence $H_\theta(O) \ge 2^N$. This concerns the explicit gap-state formulation, not a lower bound for all possible algorithms.

Obtaining $\Delta_{\min}$ for the static computation. The approximation takes a known instance θ as input. Its minimum gap can be computed on the same graph, rather than supplied by an oracle. After ordinary max-sum finds $a^\star$,

$$\max_{a \ne a^\star} \theta(a) = \max_{i \in [N]} \max_{a \in \mathcal{A}:\, a_i \ne a^\star_i} \theta(a), \qquad \Delta_{\min} = \theta(a^\star) - \max_{a \ne a^\star} \theta(a). \tag{44}$$

The identity follows from $\mathcal{A} \setminus \{a^\star\} = \bigcup_i \{a : a_i \ne a^\star_i\}$. An empty restricted domain contributes $-\infty$. Each inner problem restricts one existing domain, creates no new graphical edge, and is solved by the same max-sum decomposition. This uses at most $N + 1$ ordinary passes, with $O((N+\rho)K^{w_O+1})$ work per pass including bag-potential construction (Dechter, 1999). The certified approximation is instance dependent through $\sum_e B_e/\Delta_{\min}$; no gap-uniform polynomial-time claim follows from (43).

E Proofs for Junction-Tree Track-and-Stop

In this appendix, we prove Lemmas 7 and 8, Prop. 9, and Thm. 10. Shared estimates are collected in Sec. E.1.

Model and allocation notation. Write $\mathcal{J} = \{(e,x) : e \in [\rho],\ x \in \mathcal{A}_e\}$, $d = |\mathcal{J}|$, and $\mathcal{J}_+ = \{(e,x) \in \mathcal{J} : |\mathcal{A}_e| \geq 2\}$. The action space $\mathcal{A} = \prod_{i=1}^N \mathcal{A}_i$ is finite and $|\mathcal{A}| \geq 2$. The local means are unrestricted, i.e., $\theta \in \mathbb{R}^d$. The unique optimal action is $a^\star$, and we recall the following definitions

$$\Delta_\theta(a) = \sum_e (\theta_{e,a_e^\star} - \theta_{e,a_e}), \qquad \Delta_{\min} = \min_{a \neq a^\star} \Delta_\theta(a) > 0, \qquad \Delta_{\max} = \max_{a \neq a^\star} \Delta_\theta(a). \tag{45}$$

Fix the JT, its root, the host map $e \mapsto C(e)$, and its child-combination orders. For $\mu \in \mathcal{M}_O$, set $\mu_{e,x} = \sum_{y \in \mathcal{A}_{C(e)} : y_{\mathcal{S}_e} = x} \mu_{C(e),y} = \tilde{w}^\mu_{e,x}$. For a reference action b, define

$$D^b_\mu(a) = \sum_{e : a_e \neq b_e} \left(\mu^{-1}_{e,b_e} + \mu^{-1}_{e,a_e}\right), \qquad T_\vartheta(\mu) = \max_{a \neq a^\star_\vartheta} \frac{D^{a^\star_\vartheta}_\mu(a)}{\Delta_\vartheta(a)^2}. \tag{46}$$

The latter definition is used only when ϑ has a unique maximizer. A local numerator contribution is defined to be zero for $a_e = b_e$, not as $0 \cdot \infty$; otherwise $0^{-1} = +\infty$. Set $T^\star_\vartheta = \inf_{\mu \in \mathcal{M}_O} T_\vartheta(\mu)$ and $\|\vartheta - \theta\|_{\mathrm{loc}} = \max_{(e,x) \in \mathcal{J}} |\vartheta_{e,x} - \theta_{e,x}|$. Bag potentials $\vartheta_{C,y} = \sum_{e : C(e) = C} \vartheta_{e,y_{\mathcal{S}_e}}$ are computational aggregates, not independent statistical arms.

Experiment and timing. Let $\{X_{e,x,n} : n \geq 1\}_{(e,x) \in \mathcal{J}}$ be mutually independent streams with $X_{e,x,n} \sim \mathcal{N}(\theta_{e,x}, \sigma^2)$, for known $\sigma^2 > 0$. Playing A_s reveals the next unused sample of stream $(e, A_{s,e})$ for each e. Fresh sampling seeds are independent of all streams. With $\mathcal{F}_t$ the history of observations, actions, and consumed seeds through round t, an update using $\mathcal{F}_t$ affects A_{t+1}, not A_t. Cached targets are $\mathcal{F}_t$-measurable. In particular,

$$\begin{aligned} N_{t,e,x} &= \sum_{s=1}^t \mathbf{1}\{A_{s,e} = x\}, \qquad \hat{\theta}_{t,e,x} = \frac{1}{N_{t,e,x}} \sum_{n=1}^{N_{t,e,x}} X_{e,x,n}, \\ \bar{\mu}_{t,C,y} &= \frac{1}{t} \sum_{s=1}^t \mathbf{1}\{A_{s,C} = y\}, \quad t\bar{\mu}_{t,e,x} = N_{t,e,x}. \end{aligned} \tag{47}$$

The first $n_0 = |\mathcal{A}_0|$ rounds play a full cycle of a covering set. Hence all factor counts are positive before estimation, optimization, or stopping begins. The initial target μ^0 is induced by independent uniform agent actions. After initialization, square-numbered action rounds cycle through $\mathcal{A}_0$ and override target sampling.

Schedules and target accuracy. Let $\mathcal{U}$ be a deterministic infinite update sequence containing n_0, and write $v(t) = \max\{u \in \mathcal{U} : u \leq t\}$. For the expected stopping-time statement, assume that for some fixed $\kappa \in (0,1]$ and $t_U < \infty$,

$$v(t) \geq \kappa t \quad (t \geq t_U), \qquad \mathcal{Q} = \{q_k : k \geq 1\}, \quad q_k \uparrow \infty, \quad \frac{q_{k+1}}{q_k} \longrightarrow 1. \tag{48}$$

Here $\mathcal{Q}$ is the deterministic stopping-check sequence. Both per-round and doubling target updates satisfy the first condition; both per-round and square-time checks satisfy the second. The almost-sure target and tracking lemmas only require that $\mathcal{U}$ is infinite.

The quantized table and oracle condition are

$$\begin{aligned} &\eta_t = \frac{c_\eta}{\log(t+2)}, \quad c_\eta > 0, \qquad \hat{\theta}^{\eta_t}_{t,e,x} = \eta_t \,\mathrm{round}(\hat{\theta}_{t,e,x}/\eta_t), \\ &\mu_u \in \mathcal{M}_O, \qquad\qquad T_{\hat{\theta}^{\eta_u}_u}(\mu_u) \leq (1+\xi_u) T^\star_{\hat{\theta}^{\eta_u}_u}, \quad 0 \leq \xi_u \leq 1, \quad \xi_u \to 0. \end{aligned} \tag{49}$$

The last line applies at updates with a unique quantized maximizer. At ties the previous target is retained. Exact optimization is $\xi_u = 0$. Rounding and feasible oracle selection have fixed measurable tie rules. Such a selection exists: at a fixed time the quantized parameter takes values in a countable lattice, and one acceptable solution can be selected at each lattice point. The solver must satisfy (49); numerical termination alone is not that guarantee.

All sampling choices, schedules, and oracle tolerances are independent of δ. We define every τ_δ on the same infinite experiment. At each check, current empirical and quantized tables are recomputed even when the target is cached. If the empirical optimum is tied, set $Z_t = \underline{Z}_t = 0$. If it is unique but is not the same unique maximizer after quantization, set $\underline{Z}_t = 0$. Such checks are called inactive.

E.1 Auxiliary lemmas

Lemma 23 (Objective calculus)**.** *The minimum $T_\theta^\star$ is attained and belongs to $(0, \infty)$. The finite domain of T_θ is $\{\mu \in \mathcal{M}_O : \mu_{e,x} > 0 \; \forall (e,x) \in \mathcal{J}_+\}$. Moreover, for $L > 0$,*

$$\begin{aligned} \frac{2}{\Delta_{\max}^2} \le T_\theta^\star \le T_\theta(\mu^0) \le \frac{2d}{\Delta_{\min}^2}, \\ T_\theta(\mu) \le L \implies \min_{(e,x)\in\mathcal{J}_+} \mu_{e,x} \ge \frac{1}{L\Delta_{\max}^2}. \end{aligned} \tag{50}$$

The functional T_θ is convex. If $\mu_{e,x} \ge \lambda \nu_{e,x}$ on $\mathcal{J}_+$, for $\lambda > 0$, then $T_\theta(\mu) \le \lambda^{-1} T_\theta(\nu)$. If both collections have factor marginals at least $m > 0$, then

$$|T_\theta(\mu) - T_\theta(\nu)| \le \frac{2\rho}{m^2 \Delta_{\min}^2} \max_{(e,x)\in\mathcal{J}_+} |\mu_{e,x} - \nu_{e,x}|. \tag{51}$$

Proof. The set $\mathcal{M}_O$ is a nonempty closed subset of $\prod_C \Delta(\mathcal{A}_C)$, hence compact. Nonemptiness follows from μ^0. The extended reciprocal is convex and lower semicontinuous on $[0, \infty)$. Since each factor projection is linear, each function $D_\mu^{a^\star}(a)/\Delta_\theta(a)^2$ is convex and lower semicontinuous in μ, as is their finite maximum.

For each $(e, x) \in \mathcal{J}_+$ choose a competitor $a^{e,x}$ as follows: if $x \ne a_e^\star$, extend x to a joint action; if $x = a_e^\star$, extend any other local assignment of factor e. The Cartesian action space permits both constructions, and in either case

$$D_\mu^{a^\star}(a^{e,x}) \ge \mu_{e,x}^{-1}, \qquad T_\theta(\mu) \ge \frac{\mu_{e,x}^{-1}}{\Delta_\theta(a^{e,x})^2} \ge \frac{1}{\Delta_{\max}^2 \mu_{e,x}}. \tag{52}$$

Thus a zero active marginal gives infinite objective, and (52) also proves the sublevel bound. Conversely, positive active marginals make all finitely many reciprocals finite. Singleton factors have marginal one and do not appear in $D_\mu^{a^\star}$.

Under μ^0, $\mu^0_{e,x} = 1/|\mathcal{A}_e|$ and $D_{\mu^0}^{a^\star}(a) \le 2\sum_e |\mathcal{A}_e| = 2d$. Every competitor differs from $a^\star$ on at least one factor, since otherwise it would tie with $a^\star$. Consequently $D_\mu^{a^\star}(a) \ge 2$. These observations prove both bounds on $T_\theta^\star$ in (50).

For the domination statement, apply $\mu_{e,x}^{-1} \le \lambda^{-1}\nu_{e,x}^{-1}$ to each relevant reciprocal and then maximize over a. Finally, with $z = \max_{(e,x)\in\mathcal{J}_+} |\mu_{e,x} - \nu_{e,x}|$,

$$\begin{aligned} |\mu_{e,x}^{-1} - \nu_{e,x}^{-1}| = \frac{|\mu_{e,x} - \nu_{e,x}|}{\mu_{e,x}\nu_{e,x}} \le \frac{z}{m^2}, \\ |T_\theta(\mu) - T_\theta(\nu)| \le \max_{a\ne a^\star} \frac{|D_\mu^{a^\star}(a) - D_\nu^{a^\star}(a)|}{\Delta_\theta(a)^2} \le \frac{2\rho z}{m^2 \Delta_{\min}^2}. \end{aligned}$$

□

Lemma 24 (Uniform perturbation of the allocation problem)**.** *If $r = 2\rho\|\vartheta - \theta\|_{\mathrm{loc}}/\Delta_{\min} < 1$, then $a_\vartheta^\star = a^\star$ and, for every $\mu \in \mathcal{M}_O$,*

$$\frac{T_\theta(\mu)}{(1+r)^2} \le T_\vartheta(\mu) \le \frac{T_\theta(\mu)}{(1-r)^2}, \qquad \frac{T_\theta^\star}{(1+r)^2} \le T_\vartheta^\star \le \frac{T_\theta^\star}{(1-r)^2}. \tag{53}$$

In particular, $T_\vartheta(\mu) \le (1+\xi)T_\vartheta^\star$ implies

$$T_\theta^\star \le T_\theta(\mu) \le (1+\xi)\left(\frac{1+r}{1-r}\right)^2 T_\theta^\star. \tag{54}$$

For $\|\vartheta - \theta\|_{\mathrm{loc}} \le b$, the same conclusions for $Q_\eta(\vartheta)$ hold with $r = (2\rho b + \rho\eta)/\Delta_{\min} < 1$.

Proof. For every $a \neq a^\star$,

$$\begin{aligned} |[\vartheta(a^\star) - \vartheta(a)] - \Delta_\theta(a)| &\le \sum_e \big(|\vartheta_{e,a_e^\star} - \theta_{e,a_e^\star}| + |\vartheta_{e,a_e} - \theta_{e,a_e}|\big) \\ &\le 2\rho\|\vartheta - \theta\|_{\mathrm{loc}} = r\Delta_{\min} \le r\Delta_\theta(a). \end{aligned} \tag{55}$$

It follows that $(1-r)\Delta_\theta(a) \le \vartheta(a^\star) - \vartheta(a) \le (1+r)\Delta_\theta(a)$. The left side is positive, so $a^\star$ is still the unique maximizer and these differences equal $\Delta_\vartheta(a)$. The reference action in the two numerator functions is therefore identical. Taking reciprocal squares in this inequality gives

$$\frac{D_\mu^{a^\star}(a)}{(1+r)^2\Delta_\theta(a)^2} \le \frac{D_\mu^{a^\star}(a)}{\Delta_\vartheta(a)^2} \le \frac{D_\mu^{a^\star}(a)}{(1-r)^2\Delta_\theta(a)^2}.$$

Taking maxima over a and then minima over μ proves (53), including its extended-valued interpretation. The transfer bound follows by the explicit chain

$$\begin{aligned} T_\theta(\mu) &\le (1+r)^2 T_\vartheta(\mu) \le (1+r)^2(1+\xi)T_\vartheta^\star \\ &\le (1+\xi)\left(\frac{1+r}{1-r}\right)^2 T_\theta^\star. \end{aligned}$$

Finally, $\|Q_\eta(\vartheta) - \theta\|_{\mathrm{loc}} \le b + \eta/2$. Substitution of this upper bound in (55) yield the result. □

Lemma 25 (Exploration and adaptive Gaussian concentration)**.** *A covering set with $n_0 \le d$ exists. If R_t counts initialization and forced rounds, then $R_t \le n_0 + \lfloor\sqrt{t}\rfloor$. For $c_0 = 1/(2n_0)$ there is a deterministic t_E such that $N_{t,e,x} \ge c_0\sqrt{t}$ for all (e,x) and $t \ge t_E$. For $r > 0$, let $a_r = r^2/(2\sigma^2)$ and $A_r = 2d/(1-e^{-a_r})$. Then*

$$\mathbb{P}_\theta\big(\|\hat\theta_t - \theta\|_{\mathrm{loc}} > r\big) \le A_r e^{-a_r c_0\sqrt{t}} \quad (t \ge t_E), \qquad \hat\theta_t \longrightarrow \theta \quad \textit{a.s.} \tag{56}$$

For $h(T) = \lceil\sqrt{T}\rceil$ and $\mathcal{G}_T^{\mathrm{est}}(r) = \bigcap_{u=h(T)}^T \{\|\hat\theta_u - \theta\|_{\mathrm{loc}} \le r\}$, sufficiently large T satisfy

$$\mathbb{P}_\theta\big((\mathcal{G}_T^{\mathrm{est}}(r))^c\big) \le A_r T e^{-a_r c_0 T^{1/4}}. \tag{57}$$

Proof. Extend every (e,x) to one joint action and discard duplicates; at most d actions remain. A full initialization cycle observes each local action once. The number of later square rounds by $t \ge n_0$ is $m_t = \lfloor\sqrt{t}\rfloor - \lfloor\sqrt{n_0}\rfloor$, and every block of n_0 such rounds is a covering cycle. Thus

$$N_{t,e,x} \ge 1 + \left\lfloor \frac{m_t}{n_0} \right\rfloor \ge \frac{m_t}{n_0} \ge \frac{\sqrt{t} - 1 - \sqrt{n_0}}{n_0} \ge c_0\sqrt{t} \quad \big(t \ge 4(1+\sqrt{n_0})^2\big). \tag{58}$$

Take t_E to be an integer at least this last threshold and n_0. Counting the square rounds also gives the bound on R_t.

Let $\bar X_{e,x,n} = n^{-1}\sum_{j=1}^n X_{e,x,j}$. For deterministic n, the Gaussian moment-generating function gives, for $\lambda > 0$,

$$\begin{aligned} \mathbb{P}_\theta(\bar X_{e,x,n} - \theta_{e,x} \ge r) &\le e^{-\lambda n r}\mathbb{E}_\theta e^{\lambda\sum_{j=1}^n (X_{e,x,j} - \theta_{e,x})} \\ &= \exp(-\lambda n r + n\sigma^2\lambda^2/2) \le e^{-n a_r}, \end{aligned} \tag{59}$$

where the last line uses $\lambda = r/\sigma^2$. Applying it also to the negative sum yields $\mathbb{P}(|\bar X_{e,x,n} - \theta_{e,x}| > r) \le 2e^{-n a_r}$.

The adaptive count must not be conditioned on. Instead, the pathwise inclusion

$$\{|\hat\theta_{t,e,x} - \theta_{e,x}| > r\} \subseteq \bigcup_{n \ge \lceil c_0\sqrt{t}\rceil} \{|\bar X_{e,x,n} - \theta_{e,x}| > r\}$$

combined with (59) gives

$$\begin{aligned} \mathbb{P}_\theta(\|\hat\theta_t - \theta\|_{\mathrm{loc}} > r) &\le \sum_{(e,x)\in\mathcal{J}} 2 \sum_{n \ge \lceil c_0\sqrt{t}\rceil} e^{-n a_r} \\ &= \frac{2d\, e^{-a_r\lceil c_0\sqrt{t}\rceil}}{1 - e^{-a_r}} \le A_r e^{-a_r c_0\sqrt{t}}. \end{aligned}$$

The series over t is finite. For every integer $k \geq 1$, the first Borel-Cantelli lemma applied with $r = 1/k$ therefore gives only finitely many deviations larger than $1/k$. A countable intersection proves the almost-sure claim. Finally,

$$\mathbb{P}_\theta((\mathcal{G}_T^{\mathrm{est}}(r))^c) \leq \sum_{u=h(T)}^{T} A_r e^{-a_r c_0 \sqrt{u}} \leq A_r T e^{-a_r c_0 T^{1/4}},$$

which proves (57). This stream-indexed union argument is also used for adaptive counts in Vannella et al. (2023a, Appendix D.2), while here it is applied directly to factor-local observations. □

Lemma 26 (Bounded martingale differences)**.** *Let V_s be $\mathcal{F}_s$-measurable, $|V_s| \leq 1$, and $\mathbb{E}[V_s \mid \mathcal{F}_{s-1}] = 0$. For $M_t = \sum_{s=1}^{t} V_s$, $z > 0$, and $n \geq 1$,*

$$\mathbb{P}(|M_t| \geq tz) \leq 2e^{-tz^2/2}, \qquad \mathbb{P}(\exists t \geq n : |M_t| \geq tz) \leq \frac{2e^{-nz^2/2}}{1 - e^{-z^2/2}}, \qquad \frac{M_t}{t} \longrightarrow 0 \quad \textit{a.s.} \tag{60}$$

Proof. This is the bounded martingale-difference specialization of Azuma (1967, Sec. 2, Lem. 1 and Remark 1). We include the exponential argument. For every $v \in [-1, 1]$, convexity of the exponential gives $e^{\lambda v} \leq (1+v)e^{\lambda}/2 + (1-v)e^{-\lambda}/2$. Consequently,

$$\mathbb{E}[e^{\lambda V_s} \mid \mathcal{F}_{s-1}] \leq \cosh \lambda \leq e^{\lambda^2/2},$$
$$\mathbb{E}[e^{\lambda M_t}] = \mathbb{E}\big[e^{\lambda M_{t-1}} \mathbb{E}[e^{\lambda V_t} \mid \mathcal{F}_{t-1}]\big] \leq e^{\lambda^2/2} \mathbb{E}[e^{\lambda M_{t-1}}] \leq e^{t\lambda^2/2}.$$

The bound $\cosh \lambda \leq e^{\lambda^2/2}$ follows by integrating $(\log \cosh \lambda)' = \tanh \lambda \leq \lambda$ for $\lambda \geq 0$ and using symmetry. Markov's inequality with $\lambda = z$ yields $\mathbb{P}(M_t \geq tz) \leq \exp(-tz^2 + tz^2/2)$; the same argument for $-M_t$ proves the first inequality. Summing it over $t \geq n$ proves the second. Applying Borel-Cantelli for $z = 1/k$, $k \geq 1$, proves the last statement. □

External deviation bound. We use the general exponential-family calibration of Kaufmann and Koolen (2021, Thm. 7, equations (9)-(10)). Let $h(v) = v - \log v$ for $v \geq 1$, and let $h^{-1} : [1, \infty) \to [1, \infty)$ be its inverse. For $1 < z < e$ and $y \geq 0$, define

$$\widetilde{h}_z(y) = \begin{cases} e^{1/h^{-1}(y)} h^{-1}(y), & y \geq h(1/\log z), \\ z(y - \log\log z), & y < h(1/\log z), \end{cases} \qquad \mathcal{C}_{\exp}(x) = 2\widetilde{h}_{3/2}\left(\frac{h^{-1}(1+x) + \log(2\zeta(2))}{2}\right). \tag{61}$$

Here $\zeta(2) = \sum_{n \geq 1} n^{-2}$; the inverse in the first branch is used only on its domain.

Lemma 27 (Self-normalized bound: external input)**.** *For an adaptive sampling process on d independent Gaussian arms of known variance σ^2, and any $x > 0$, Kaufmann and Koolen (2021, Thm. 7) implies*

$$\mathbb{P}\left(\exists n : \sum_{j=1}^{d} N_j(n) \frac{(\hat{\theta}_j(n) - \theta_j)^2}{2\sigma^2} > 3 \sum_{j=1}^{d} \log(1 + \log N_j(n)) + d\mathcal{C}_{\exp}(x/d)\right) \leq e^{-x}. \tag{62}$$

Proof. For Gaussian arms, $\mathrm{KL}(\mathcal{N}(m, \sigma^2), \mathcal{N}(m', \sigma^2)) = (m - m')^2/(2\sigma^2)$. Thm. 7 of Kaufmann and Koolen (2021) bounds the sum of the positive parts of the corresponding centered deviations. The inclusion follows from $\sum_j u_j \leq \sum_j [u_j]_+$ with $u_j = N_j(n)(\hat{\theta}_j(n) - \theta_j)^2/(2\sigma^2) - 3\log(1 + \log N_j(n))$. Its application to our simultaneous observations is justified separately in the proof of Thm. 10. □

Lemma 28 (Boundary envelopes)**.** *Set $L_\delta = \log(1/\delta)$, $a_\delta = d\mathcal{C}_{\exp}(L_\delta/d)$, and*

$$\beta(t, \delta) = a_\delta + 3 \sum_{(e,x) \in \mathcal{J}} \log(1 + \log N_{t,e,x}), \qquad g(t) = 3d \log(1 + \log t). \tag{63}$$

For $t \geq n_0$, $L_\delta \leq a_\delta \leq \beta(t, \delta) \leq a_\delta + g(t)$. Also $a_\delta / L_\delta \to 1$ as $\delta \downarrow 0$, $g(t) = o(t)$, and, for every $k > 0$, $(a_\delta + g(kL_\delta))/L_\delta \to 1$. If $a_\delta > 3d$, then $(a_\delta + g(t))/t$ is decreasing on $[1, \infty)$.

Proof. From $h(v) \le v$ and $h(v)/v \to 1$, we have $h^{-1}(y) \ge y$ and $h^{-1}(y)/y \to 1$. Both branches of $\widetilde{h}_{3/2}$ are at least their nonnegative argument; its large-argument branch is asymptotic to that argument. Therefore

$$\mathcal{C}_{\exp}(x) \ge h^{-1}(1+x) + \log(2\zeta(2)) \ge x, \qquad \frac{\mathcal{C}_{\exp}(x)}{x} \longrightarrow 1.$$

Together with $1 \le N_{t,e,x} \le t$, these prove the envelopes and limits. For monotonicity,

$$\frac{d}{dt}\frac{a_\delta + g(t)}{t} = \frac{3d/(1+\log t) - a_\delta - g(t)}{t^2} < 0 \quad (a_\delta > 3d,\ t \ge 1). \tag{64}$$

The boundary (63) is the specialization of Kaufmann and Koolen (2021, Prop. 15, equation (20)). Its lower envelope will be needed for the almost-sure lower limit, and its monotone upper envelope for the expectation proof. □

E.2 Proof of Lem. 7: consistency of the JT target

Proof of Lem. 7. Let $\Omega_{\text{est}} = \{\|\hat{\theta}_t - \theta\|_{\text{loc}} \to 0\}$. By Lem. 25, $\mathbb{P}_\theta(\Omega_{\text{est}}) = 1$. On this event, at every scheduled update define

$$b_u = \|\hat{\theta}_u - \theta\|_{\text{loc}}, \qquad r_u = \frac{2\rho b_u + \rho\eta_u}{\Delta_{\min}}. \tag{65}$$

Since $b_u \to 0$ and $\eta_u \to 0$, $r_u \to 0$. In particular, there exists an almost-surely finite $u_\star$ such that $r_u < 1$ for every scheduled $u \ge u_\star$. Lem. 24 gives

$$\hat{\theta}_u^{\eta_u}(a^\star) - \hat{\theta}_u^{\eta_u}(a) \ge (1 - r_u)\Delta_\theta(a) > 0 \quad \text{for every } a \ne a^\star. \tag{66}$$

Thus every scheduled update from that point succeeds. The value and oracle-output bounds become

$$\begin{aligned} \frac{T_\theta^\star}{(1+r_u)^2} \le T^\star_{\hat{\theta}_u^{\eta_u}} \le \frac{T_\theta^\star}{(1-r_u)^2}, \\ T_\theta^\star \le T_\theta(\mu_u) \le (1+\xi_u)\left(\frac{1+r_u}{1-r_u}\right)^2 T_\theta^\star. \end{aligned} \tag{67}$$

Letting $u \to \infty$ along $\mathcal{U}$ proves both conclusions by the squeeze theorem. This also proves the statement for a feasible approximate solver satisfying (49).

Let p_s denote the target cached before action round s, and let $j(s) < s$ be the update time at which it was last successfully computed. Once (66) holds, $j(s) \to \infty$ because $\mathcal{U}$ is infinite. Hence

$$T_\theta(p_s) = T_\theta(\mu_{j(s)}) \longrightarrow T_\theta^\star. \tag{68}$$

This step only uses infinitude of $\mathcal{U}$. The stronger schedule condition (48) will enter the expectation proof. □

Uniform version used later. The preceding proof has the following deterministic consequence: for any $b, \eta, \xi \ge 0$ with $(2\rho b + \rho\eta)/\Delta_{\min} < 1$, every feasible $(1+\xi)$-optimal oracle output at a table within local distance b of θ satisfies (54). Thus the late-time approximation bound is uniform over histories and over all acceptable oracle outputs.

For completeness, objective convergence also yields convergence to the set of optimizers. Let $\mathcal{M}_\theta^\star = \arg\min_{\mu \in \mathcal{M}_O} T_\theta(\mu)$. It is nonempty and compact by Lem. 23. If a sequence of targets stayed a positive distance from this set, compactness would supply $\mu_{u_k} \to \mu_\infty$ with the same positive distance, while lower semicontinuity would give

$$T_\theta(\mu_\infty) \le \liminf_{k\to\infty} T_\theta(\mu_{u_k}) = T_\theta^\star,$$

a contradiction. Therefore $\operatorname{dist}(\mu_u, \mathcal{M}_\theta^\star) \to 0$ a.s.

E.3 Proof of Lem. 8: tracking the realized JT allocation

The proof separates the conditional law of an ancestral sample, concentration around its predictable marginals, and the transfer of target optimality to the observed allocation. Note that deterministic tracking results, such as Garivier and Kaufmann (2016, Appendix B, Lemmas 7-8) and Vannella et al. (2023a, Appendix C.2), do not establish the required statement for this randomized sampler.

Lemma 29 (Zero-safe JT sampling)**.** *For every $\mu \in \mathcal{M}_O$, the rooted ancestral sampler produces a law P_μ on $\mathcal{A}$ such that $P_\mu(A_C = y) = \mu_{C,y}$ for every bag C and $y \in \mathcal{A}_C$.*

Proof. For the root R, draw $A_R \sim \mu_R$. For a child C with parent $P(C)$ and separator $S_C = C \cap P(C)$, choose a conditional kernel

$$K_C(v \mid z) = \begin{cases} \mu_{C,(z,v)}/\mu_{S_C,z}, & \mu_{S_C,z} > 0, \\ \pi_{C,z}(v), & \mu_{S_C,z} = 0, \end{cases} \quad v \in \mathcal{A}_{C \setminus S_C},\ z \in \mathcal{A}_{S_C}, \tag{69}$$

where $\pi_{C,z}$ is any fixed probability distribution. Nonnegativity and marginalization give

$$\sum_v \mu_{C,(z,v)} = \mu_{S_C,z}, \qquad \mu_{S_C,z} = 0 \implies \mu_{C,(z,v)} = 0 \quad \forall v. \tag{70}$$

Hence K_C is normalized for every z, including zero-mass states.

Running intersection implies that $C \cap V_{\text{past}} = S_C$ when C is processed in a root-to-leaf order, where V_{past} denotes previously assigned variables. Indeed, a variable shared by C and an earlier bag must occur at every bag on their path, in particular at $P(C)$. The sets R and $C \setminus S_C$, over nonroot bags, therefore partition the agent variables covered by the tree. The ancestral construction defines the normalized joint law

$$P_\mu(a) = \mu_{R,a_R} \prod_{C \neq R} K_C(a_{C \setminus S_C} \mid a_{S_C}). \tag{71}$$

Variables absent from every factor may be appended independently.

We prove correctness of the bag marginals by induction along the traversal. The root is immediate. Suppose the parent $P(C)$ has marginal $\mu_{P(C)}$. Separator consistency implies $P_\mu(A_{S_C} = z) = \mu_{S_C,z}$. For every $y = (z, v)$,

$$\begin{aligned} P_\mu(A_C = y) &= P_\mu(A_{S_C} = z) K_C(v \mid z) \\ &= \begin{cases} \mu_{S_C,z}\, \mu_{C,y}/\mu_{S_C,z} = \mu_{C,y}, & \mu_{S_C,z} > 0, \\ 0 = \mu_{C,y}, & \mu_{S_C,z} = 0. \end{cases} \end{aligned} \tag{72}$$

Later conditional draws are normalized and do not overwrite assigned variables, so they preserve these marginals. Induction proves the assertion. This is the constructive junction-tree extension underlying local-to-global consistency (see Wainwright and Jordan (2008, Sec. 2.5.2)). □

Let $\mathcal{I}$ be the deterministic set of target-sampling rounds, excluding initialization and forced exploration. Define the predictable bag allocation

$$q_s = \begin{cases} p_s, & s \in \mathcal{I}, \\ \delta^{a_s^{\exp}}, & s \notin \mathcal{I}, \end{cases} \qquad \bar{q}_t = \frac{1}{t} \sum_{s=1}^{t} q_s, \tag{73}$$

where $a_s^{\exp}$ is the specified initialization or covering-set action and $\delta^a_{C,y} = \mathbf{1}\{a_C = y\}$. q_s belongs to $\mathcal{M}_O$ and is $\mathcal{F}_{s-1}$-measurable.

Lemma 30 (Bag and factor martingale tracking)**.** *Write $B_O = \sum_{C \in \mathcal{C}_O} |\mathcal{A}_C|$. For every $t, n \geq 1$ and $z > 0$,*

$$\begin{aligned} \mathbb{P}_\theta(\|\bar{\mu}_t - \bar{q}_t\|_\infty \geq z) &\leq 2B_O e^{-tz^2/2}, \\ \mathbb{P}_\theta(\exists t \geq n : \|\bar{\mu}_t - \bar{q}_t\|_\infty \geq z) &\leq \frac{2B_O e^{-nz^2/2}}{1 - e^{-z^2/2}}. \end{aligned} \tag{74}$$

Replacing the bag norm by $\max_{(e,x) \in \mathcal{J}} |\bar{\mu}_{t,e,x} - \bar{q}_{t,e,x}|$ gives the same inequalities with d in place of B_O. Both differences tend to zero almost surely.

Proof. Fix C and $y \in \mathcal{A}_C$. Conditional on $\mathcal{F}_{s-1}$, the cached target p_s is fixed and the round-s random seed is fresh. Lem. 29 gives the conditional marginal on a target-sampling round. At all other rounds the action is already determined by the past. Therefore

$$\begin{aligned}\mathbb{E}_\theta[\mathbf{1}\{A_{s,C} = y\} \mid \mathcal{F}_{s-1}] &= \mathbf{1}\{s \in \mathcal{I}\}p_{s,C,y} + \mathbf{1}\{s \notin \mathcal{I}\}\mathbf{1}\{a^{\exp}_{s,C} = y\} \\ &= q_{s,C,y}.\end{aligned} \tag{75}$$

Define

$$V_{s,C,y} = \mathbf{1}\{A_{s,C} = y\} - q_{s,C,y}, \qquad M_{t,C,y} = \sum_{s=1}^{t} V_{s,C,y}.$$

The variable $V_{s,C,y}$ is $\mathcal{F}_s$-measurable, with $|V_{s,C,y}| \le 1$ and $\mathbb{E}[V_{s,C,y} \mid \mathcal{F}_{s-1}] = 0$. Hence $M_{t,C,y}$ is an integrable martingale and

$$\frac{M_{t,C,y}}{t} = \bar{\mu}_{t,C,y} - \bar{q}_{t,C,y}. \tag{76}$$

Applying Lem. 26 to each coordinate and taking a finite union gives the first line of (74). Summing the first line over $t \ge n$ gives the second. For $z = 1/k$, the right side tends to zero as $n \to \infty$. A countable intersection over k proves simultaneous almost-sure convergence.

To obtain the factor-coordinate bound, sum (75) over the host-bag assignments restricting to x:

$$\mathbb{E}_\theta[\mathbf{1}\{A_{s,e} = x\} \mid \mathcal{F}_{s-1}] = \sum_{y : y_{\mathcal{S}_e} = x} q_{s,C(e),y} = q_{s,e,x}.$$

The martingale $\sum_{s \le t}(\mathbf{1}\{A_{s,e} = x\} - q_{s,e,x})$ again has increments bounded by one. Applying Lem. 26 and a union bound over d coordinates proves the claim. Neither independence between bag coordinates nor conditioning on the future sequence of targets is used. □

Proof of Lem. 8. Work on the probability-one intersection of target consistency and Lem. 30. Given $\epsilon > 0$, (68) supplies a finite s_ϵ such that $T_\theta(p_s) \le T^\star_\theta + \epsilon$ for every target-sampling round $s \ge s_\epsilon$. Put

$$\mathcal{L}_t = \mathcal{I} \cap \{s_\epsilon, \dots, t\}, \quad m_t = |\mathcal{L}_t|, \quad \nu_t = \frac{1}{m_t}\sum_{s \in \mathcal{L}_t} p_s.$$

These are defined for all sufficiently large t, since $m_t \ge t - s_\epsilon - R_t$ and $R_t/t \to 0$. Hence $m_t/t \to 1$. Convexity and coordinatewise nonnegativity give

$$\begin{aligned}T_\theta(\nu_t) &\le \frac{1}{m_t}\sum_{s \in \mathcal{L}_t} T_\theta(p_s) \le T^\star_\theta + \epsilon, \\ \bar{q}_{t,e,x} &= \frac{1}{t}\sum_{s=1}^{t} q_{s,e,x} \ge \frac{1}{t}\sum_{s \in \mathcal{L}_t} p_{s,e,x} = \frac{m_t}{t}\nu_{t,e,x}, \\ T^\star_\theta &\le T_\theta(\bar{q}_t) \le \frac{t}{m_t} T_\theta(\nu_t) \le \frac{t}{m_t}(T^\star_\theta + \epsilon).\end{aligned} \tag{77}$$

The middle inequality in the last line is the domination statement of Lem. 23. Let $t \to \infty$ and then $\epsilon \downarrow 0$ to obtain $T_\theta(\bar{q}_t) \to T^\star_\theta$. Restricting Jensen's inequality to late targets avoids assigning finite objective values to exceptional early targets or forced point masses.

Eventually $T_\theta(\bar{q}_t) \le 2T^\star_\theta$, so (50) gives

$$\bar{q}_{t,e,x} \ge m_\theta := \frac{1}{2T^\star_\theta \Delta^2_{\max}} \quad ((e,x) \in \mathcal{J}_+).$$

Set $z_t = \max_{\mathcal{J}} |\bar{\mu}_{t,e,x} - \bar{q}_{t,e,x}|$. Lem. 30 implies $z_t \to 0$, hence eventually $\bar{\mu}_{t,e,x} \ge m_\theta/2$ on $\mathcal{J}_+$. Applying (51),

$$|T_\theta(\bar{\mu}_t) - T_\theta(\bar{q}_t)| \le \frac{8\rho}{m^2_\theta \Delta^2_{\min}} z_t \longrightarrow 0. \tag{78}$$

Together with (77), this proves the lemma. □

Consequence for the unquantized empirical objective. Let $r_t = 2\rho\|\hat{\theta}_t - \theta\|_{\mathrm{loc}}/\Delta_{\min}$. Eventually $r_t < 1$ and the empirical best action is $a^\star$. Lem. 24 yields

$$\frac{T_\theta(\bar{\mu}_t)}{(1+r_t)^2} \le T_{\hat{\theta}_t}(\bar{\mu}_t) \le \frac{T_\theta(\bar{\mu}_t)}{(1-r_t)^2} \quad \Longrightarrow \quad T_{\hat{\theta}_t}(\bar{\mu}_t) \longrightarrow T_\theta^\star \quad \text{a.s.} \tag{79}$$

This assertion involves realized observations, not the quantized target objective.

E.4 Proof of Prop. 9: conservative JT evidence

E.4.1 The adaptive Gaussian likelihood and exact GLR

Fix $t \ge n_0$. Denote by $K_s(a \mid H_{s-1})$ the prescribed, possibly randomized, action-selection kernel given the observed history. The likelihood under a candidate local-mean vector λ factors as

$$\mathcal{L}_t(\lambda) = \prod_{s=1}^{t} K_s(A_s \mid H_{s-1}) \prod_{e=1}^{\rho} \varphi_{\sigma^2}(Y_{s,e} - \lambda_{e,A_{s,e}}), \tag{80}$$

where φ_{σ^2} is the centered Gaussian density. For a fixed observed history, the K_s factors do not depend on λ and cancel in likelihood ratios. Regrouping by local arm and completing squares gives

$$\begin{aligned} \log \frac{\mathcal{L}_t(\hat{\theta}_t)}{\mathcal{L}_t(\lambda)} &= \frac{1}{2\sigma^2} \sum_{e,x} \sum_{n=1}^{N_{t,e,x}} \left[(X_{e,x,n} - \lambda_{e,x})^2 - (X_{e,x,n} - \hat{\theta}_{t,e,x})^2 \right] \\ &= \frac{1}{2\sigma^2} \sum_{e,x} N_{t,e,x} (\hat{\theta}_{t,e,x} - \lambda_{e,x})^2 =: \mathcal{D}_t(\lambda). \end{aligned} \tag{81}$$

Thus adaptive selection does not alter the pathwise likelihood calculation. This is the Gaussian specialization of the generalized likelihood construction in Kaufmann and Koolen (2021, Sec. 4.1).

For a unique empirical maximizer b, let $\mathcal{B}_{b,a} = \{\lambda : \lambda(a) \ge \lambda(b)\}$, $a \ne b$, and $\mathcal{B}_b = \bigcup_{a \ne b} \mathcal{B}_{b,a}$. The following projection derives the statistic rather than assuming its form.

Lemma 31 (Pairwise projection)**.** *For every $a \ne b$, with $\hat{\Delta}_t^b(a) = \hat{\theta}_t(b) - \hat{\theta}_t(a) > 0$,*

$$\inf_{\lambda \in \mathcal{B}_{b,a}} \mathcal{D}_t(\lambda) = \frac{\hat{\Delta}_t^b(a)^2}{2\sigma^2 \sum_{e: a_e \ne b_e} (N_{t,e,b_e}^{-1} + N_{t,e,a_e}^{-1})}. \tag{82}$$

Proof. Let $J_b(a) = \{e : a_e \ne b_e\}$. Only the coordinates $(e, b_e), (e, a_e)$, $e \in J_b(a)$, enter the half-space constraint. Every other coordinate can be kept at its empirical mean. On the relevant coordinates index by j, let $n_j > 0$ be the count, and set $v_j = 1$ at (e, b_e) and $v_j = -1$ at (e, a_e). For $h_j = \lambda_j - \hat{\theta}_j$ and $\hat{\Delta} = \hat{\Delta}_t^b(a)$, the constraint is $\sum_j v_j h_j \le -\hat{\Delta}$. Set

$$R = \sum_j \frac{v_j^2}{n_j} = \sum_{e \in J_b(a)} (N_{t,e,b_e}^{-1} + N_{t,e,a_e}^{-1}).$$

For every feasible h, weighted Cauchy-Schwarz gives

$$\hat{\Delta}^2 \le \left(\sum_j v_j h_j \right)^2 = \left(\sum_j \frac{v_j}{\sqrt{n_j}} \sqrt{n_j} h_j \right)^2 \le R \sum_j n_j h_j^2. \tag{83}$$

The choice $h_j^\star = -\hat{\Delta} v_j / (n_j R)$ is feasible with equality, and

$$\sum_j v_j h_j^\star = -\hat{\Delta}, \qquad \sum_j n_j (h_j^\star)^2 = \frac{\hat{\Delta}^2}{R^2} \sum_j \frac{v_j^2}{n_j} = \frac{\hat{\Delta}^2}{R}.$$

It attains (83). Division by $2\sigma^2$ proves (82). □

Since $\mathcal{B}_b$ is a finite union and $N_{t,e,x} = t\bar{\mu}_{t,e,x}$,

$$\begin{aligned} Z_t &= \inf_{\lambda \in \mathcal{B}_b} \mathcal{D}_t(\lambda) = \min_{a \neq b} \inf_{\lambda \in \mathcal{B}_{b,a}} \mathcal{D}_t(\lambda) \\ &= \frac{t}{2\sigma^2} \min_{a \neq b} \frac{\hat{\Delta}_t^b(a)^2}{D^b_{\bar{\mu}_t}(a)} = \frac{t}{2\sigma^2 T_{\hat{\theta}_t}(\bar{\mu}_t)}. \end{aligned} \tag{84}$$

At an empirical tie, $\hat{\theta}_t$ lies in an alternative half-space and the likelihood distance is zero, agreeing with $Z_t = 0$.

E.4.2 Exact computation of the quantized gap profile

At an active check, b is the common unique best action of $\hat{\theta}_t$ and $\hat{\theta}_t^{\eta_t}$. Define the local quantities

$$g_{e,x} = \hat{\theta}^{\eta_t}_{t,e,b_e} - \hat{\theta}^{\eta_t}_{t,e,x}, \qquad c_{e,x} = \begin{cases} 0, & x = b_e, \\ \bar{\mu}^{-1}_{t,e,b_e} + \bar{\mu}^{-1}_{t,e,x}, & x \neq b_e. \end{cases} \tag{85}$$

All costs are finite after initialization. If $\mathcal{E}_C = \{e : C(e) = C\}$, set $g_{C,y} = \sum_{e \in \mathcal{E}_C} g_{e,y_{S_e}}$ and $c_{C,y} = \sum_{e \in \mathcal{E}_C} c_{e,y_{S_e}}$. For the rooted subtree $\mathcal{T}_C$, write $V_C = \bigcup_{D \in \mathcal{T}_C} D$ and $\mathcal{F}_C = \bigcup_{D \in \mathcal{T}_C} \mathcal{E}_D$. For $C \neq R$, with separator S_C, define the feasible subtree completions and message by

$$\begin{aligned} \mathcal{X}_C(z,s) &= \left\{ a_{V_C} \in \mathcal{A}_{V_C} : a_{S_C} = z,\ \sum_{e \in \mathcal{F}_C} g_{e,a_e} = s \right\}, \\ M_C(z,s) &= \max_{a_{V_C} \in \mathcal{X}_C(z,s)} \sum_{e \in \mathcal{F}_C} c_{e,a_e}, \qquad \max \varnothing = -\infty. \end{aligned} \tag{86}$$

At the root omit z and its constraint. Intermediate gaps s may be negative. The following recurrence retains all attainable values.

Lemma 32 (Gap-profile recursion)**.** *For $C \neq R$,*

$$M_C(z,s) = \max_{\substack{y \in \mathcal{A}_C,\ (s_D)_{D \in \mathrm{ch}(C)} \\ y_{S_C} = z,\ g_{C,y} + \sum_D s_D = s}} \left[c_{C,y} + \sum_{D \in \mathrm{ch}(C)} M_D(y_{C \cap D}, s_D) \right]. \tag{87}$$

The root version satisfies $M_R(s) = \Phi_{\hat{\theta}_t^{\eta_t}, \bar{\mu}_t}(s)$ for every attainable $s > 0$.

Proof. For a leaf, $V_C = C$ and $\mathcal{F}_C = \mathcal{E}_C$, so (87) is exactly (86). Assume the child statements hold. Every $a_{V_C} \in \mathcal{X}_C(z,s)$ induces $y = a_C$, child gaps $s_D = \sum_{e \in \mathcal{F}_D} g_{e,a_e}$, and compatible child completions. Since the hosted factor sets are disjoint,

$$\begin{aligned} s &= g_{C,y} + \sum_D s_D, \\ \sum_{e \in \mathcal{F}_C} c_{e,a_e} &= c_{C,y} + \sum_D \sum_{e \in \mathcal{F}_D} c_{e,a_e} \leq c_{C,y} + \sum_D M_D(y_{C \cap D}, s_D). \end{aligned}$$

Maximizing shows that the left side of (87) is no greater than the right side.

Conversely, fix a finite-valued candidate on its right. Each child maximum is attained because its completion set is finite. Choose a maximizing completion $a^D_{V_D}$ for every child. Running intersection gives $V_D \cap V_{D'} \subseteq C$ for distinct children and $V_D \cap C = C \cap D$. Thus all chosen completions agree on their common variables through y. They concatenate into one $a_{V_C} \in \mathcal{X}_C(z,s)$ whose gap and cost equal the candidate's sums. This proves the reverse inequality. Infeasible candidates contribute $-\infty$ and do not affect the identity.

At the root all factors are counted once, so its gap is $\hat{\theta}_t^{\eta_t}(b) - \hat{\theta}_t^{\eta_t}(a)$ and its cost is $D^b_{\bar{\mu}_t}(a)$. Uniqueness of b means that positive root gaps correspond precisely to competitors. Hence the claimed identity with Φ follows. The root restriction $s > 0$ is not imposed at intermediate tables. □

E.4.3 Proof of the stopping certificate

Proof of Prop. 9. At an inactive check, $\underline{Z}_t = 0 \le Z_t$, so consider an active check. For each $a \neq b$, put $s_t(a) = \hat{\theta}_t^{\eta_t}(b) - \hat{\theta}_t^{\eta_t}(a)$ and $\ell_t(a) = [s_t(a) - \rho\eta_t]_+$. Nearest-grid rounding implies

$$\begin{aligned} |s_t(a) - \hat{\Delta}_t^b(a)| &\le \sum_e \big(|\hat{\theta}^{\eta_t}_{t,e,b_e} - \hat{\theta}_{t,e,b_e}| + |\hat{\theta}^{\eta_t}_{t,e,a_e} - \hat{\theta}_{t,e,a_e}|\big) \le \rho\eta_t, \\ 0 \le \ell_t(a) \le \hat{\Delta}_t^b(a), &\qquad s_t(a) - \rho\eta_t \ge \hat{\Delta}_t^b(a) - 2\rho\eta_t. \end{aligned} \tag{88}$$

All $D^b_{\hat{\mu}_t}(a)$ are finite and positive. For a fixed bucket s, its corrected numerator is constant, so Lem. 32 gives

$$\begin{aligned} \underline{Z}_t &= \frac{t}{2\sigma^2} \min_{s>0} \frac{[s - \rho\eta_t]_+^2}{\max_{a:s_t(a)=s} D^b_{\hat{\mu}_t}(a)} \\ &= \min_{a\neq b} \frac{t\ell_t(a)^2}{2\sigma^2 D^b_{\hat{\mu}_t}(a)} \le \min_{a\neq b} \frac{t\hat{\Delta}_t^b(a)^2}{2\sigma^2 D^b_{\hat{\mu}_t}(a)} = Z_t. \end{aligned} \tag{89}$$

The bucket equality remains true if its corrected numerator is zero.

For the asymptotic reverse bound define $r_t = 2\rho\eta_t / \hat{\Delta}_{\min,t}$ at active checks. If $r_t < 1$, (88) implies

$$\ell_t(a) \ge \hat{\Delta}_t^b(a) - 2\rho\eta_t \ge (1 - r_t)\hat{\Delta}_t^b(a) > 0.$$

Combining this pointwise bound with (89) yields

$$(1 - r_t)^2 Z_t \le \underline{Z}_t \le Z_t. \tag{90}$$

Lem. 25 and $\eta_t \to 0$ make both empirical instances converge to θ. By (55), their maximizers are eventually uniquely $a^\star$, so every sufficiently late check is active. Then the finite family of empirical gaps converges uniformly, giving $\hat{\Delta}_{\min,t} \to \Delta_{\min} > 0$ and $r_t \to 0$. Since $Z_t > 0$ at active checks, division in (90) proves $\underline{Z}_t / Z_t \to 1$ along the check sequence. □

E.5 Proof of Thm. 10: correctness and asymptotic optimality

We first establish finite-time error control, then the almost-sure stopping rate, and finally the expectation bound. The latter uses a late-interval good-event argument, as in Vannella et al. (2023a, Appendix D.2) and Tzeng et al. (2023, Appendix D.3), with additional control of randomized JT sampling and quantization.

E.5.1 Serialization and finite-time correctness

Lemma 33 (Serialization of semi-bandit observations)**.** *The observations obtained in global rounds $s = 1, 2, \ldots$, revealed in the order $e = 1, \ldots, \rho$ within each round, form an adaptive sampling experiment on the d independent Gaussian streams. At micro-time $n = \rho t$, the counts and estimates equal $N_{t,e,x}$ and $\hat{\theta}_{t,e,x}$.*

Proof. Let U_s contain the independent seeds used to draw the full action A_s, so A_s is measurable with respect to $\mathcal{F}_{s-1} \vee \sigma(U_s)$. Before observing factor e of round s, use

$$\mathcal{H}_{s,e-1} = \mathcal{F}_{s-1} \vee \sigma(U_s, A_s, Y_{s,1}, \ldots, Y_{s,e-1}).$$

The selected stream $J_{\rho(s-1)+e} = (e, A_{s,e})$ is $\mathcal{H}_{s,e-1}$-measurable. Its next unused observation is $Y_{s,e} = X_{e,A_{s,e},N_{s-1,e,A_{s,e}}+1}$. Independence of unused stream entries, and independence of the seeds, imply

$$\begin{aligned} \mathbb{P}_\theta(Y_{s,e} \in B \mid \mathcal{H}_{s,e-1}) &= \int_B \varphi_{\sigma^2}(y - \theta_{e,A_{s,e}})\, dy, \\ \mathbb{E}_\theta[e^{\lambda(Y_{s,e} - \theta_{e,A_{s,e}})} \mid \mathcal{H}_{s,e-1}] &= e^{\sigma^2\lambda^2/2}. \end{aligned} \tag{91}$$

For clarity, the unused-entry assertion follows by partitioning according to the finite revealed counts at time $s-1$: each event describing the preceding history depends only on revealed stream prefixes and consumed seeds, and is independent of each next unrevealed entry. Rewards revealed earlier in the

same round belong to different factor streams and preserve this independence. This proves (91) first for cylinder events in the history and then for the full sigma-field by the monotone-class argument.

The first selection within each global round uses a fresh seed, while later within-round selections are already fixed. This is exactly the randomized adaptive sampling allowed in Kaufmann and Koolen (2021, Sec. 2.1). At micro-time ρt, counting all completed reveals gives

$$N^{\mathrm{ser}}_{\rho t,e,x} = \sum_{s\le t} \mathbf{1}\{A_{s,e} = x\} = N_{t,e,x}, \qquad \hat{\theta}^{\mathrm{ser}}_{\rho t,e,x} = \hat{\theta}_{t,e,x}.$$

Thus Lem. 27 applies at every complete global round. It is not applied to aggregated bag rewards. □

Let $\mathcal{V}_\delta = \{\exists t \ge n_0 : \mathcal{D}_t(\theta) > \beta(t,\delta)\}$. Lemams 27 and 33, with $x = L_\delta$, give $\mathbb{P}_\theta(\mathcal{V}_\delta) \le \delta$. On an erroneous active check, $b = \hat{a}_t \ne a^\star$ and $\theta \in \mathcal{B}_b$. Hence (84) implies $Z_t \le \mathcal{D}_t(\theta)$, and the complete error inclusion is

$$\begin{aligned}\{\tau_\delta < \infty,\ \hat{a}_{\tau_\delta} \ne a^\star\} &\subseteq \bigcup_{t\in\mathcal{Q}} \{\hat{a}_t \ne a^\star,\ \underline{Z}_t > \beta(t,\delta)\} \\ &\subseteq \bigcup_{t\in\mathcal{Q}} \{\hat{a}_t \ne a^\star,\ Z_t > \beta(t,\delta)\} \\ &\subseteq \mathcal{V}_\delta.\end{aligned} \tag{92}$$

Taking probabilities proves error control without assuming finiteness of τ_δ in advance. This is the GLR confidence-region argument of Kaufmann and Koolen (2021, Prop. 15), with the additional pathwise comparison (89).

E.5.2 Almost-sure evidence growth and normalized stopping time

Set $c_\theta = (2\sigma^2 T^\star_\theta)^{-1} > 0$. Equations (79), (84), and (90) imply, on one probability-one event,

$$\frac{Z_t}{t} = \frac{1}{2\sigma^2 T_{\hat{\theta}_t}(\bar{\mu}_t)} \longrightarrow c_\theta, \qquad \frac{\underline{Z}_t}{t} = \frac{\underline{Z}_t}{Z_t}\frac{Z_t}{t} \longrightarrow c_\theta \quad (t \in \mathcal{Q}). \tag{93}$$

For each fixed δ, $\beta(t,\delta) \le a_\delta + g(t) = o(t)$, whereas the second limit is positive. Thus $\tau_\delta < \infty$ almost surely. Together with (92), this proves δ-correctness.

We next prove the normalized limit on the common infinite experiment. Let $q^+(x) = \min\{q \in \mathcal{Q} : q \ge x\}$ and $q^-(x) = \max\{q \in \mathcal{Q} : q \le x\}$, for sufficiently large x. From (48),

$$1 \le \frac{q^+(x)}{x} \le \frac{q_k}{q_{k-1}} \longrightarrow 1 \quad (q_{k-1} < x \le q_k), \qquad \frac{q^-(x)}{x} \longrightarrow 1. \tag{94}$$

Fix $\epsilon \in (0, 1/2)$. On the event of (93), there is a finite t_ϵ such that

$$(1-\epsilon)c_\theta t \le \underline{Z}_t \le (1+\epsilon)c_\theta t \quad \text{for all } t \in \mathcal{Q},\ t \ge t_\epsilon. \tag{95}$$

For the upper bound, take $x_\delta = (1+2\epsilon)L_\delta/[(1-\epsilon)c_\theta]$. Equations (94) and (95), and Lem. 28, give

$$\begin{aligned}\underline{Z}_{q^+(x_\delta)} &\ge (1-\epsilon)c_\theta q^+(x_\delta) \ge (1+2\epsilon)L_\delta, \\ \beta(q^+(x_\delta),\delta) &\le a_\delta + g(q^+(x_\delta)) = (1+o(1))L_\delta.\end{aligned}$$

Thus $\tau_\delta \le q^+(x_\delta)$ for all sufficiently small δ, and

$$\limsup_{\delta\downarrow 0} \frac{\tau_\delta}{L_\delta} \le \frac{1+2\epsilon}{(1-\epsilon)c_\theta}. \tag{96}$$

For the lower bound, let $M_\epsilon = \max\{\underline{Z}_t : t \in \mathcal{Q},\ t < t_\epsilon\}$, with $\max\varnothing = 0$. This is finite almost surely. If $L_\delta > M_\epsilon$, the lower boundary envelope excludes all these early checks. For any later check with $t \le L_\delta/[(1+2\epsilon)c_\theta]$,

$$\underline{Z}_t \le (1+\epsilon)c_\theta t \le \frac{1+\epsilon}{1+2\epsilon}L_\delta < L_\delta \le \beta(t,\delta).$$

It follows that

$$\liminf_{\delta\downarrow 0} \frac{\tau_\delta}{L_\delta} \ge \frac{1}{(1+2\epsilon)c_\theta}. \tag{97}$$

Sending $\epsilon \downarrow 0$ in (96)-(97) proves the stronger common-path statement

$$\mathbb{P}_\theta\left(\lim_{\delta\downarrow 0}\frac{\tau_\delta}{\log(1/\delta)} = 2\sigma^2 T_\theta^\star\right) = 1. \tag{98}$$

In particular, it implies the almost-sure limsup stated in Thm. 10 and convergence in probability. The lower half uses the lower envelope of the stopping boundary, not an expected-sample lower bound.

E.5.3 Quantitative late-interval estimates

The expected stopping-time bound requires control uniform over late horizons. We derive it with explicit approximation, sampling, and exploration losses. Fix $\epsilon \in (0, 1/2)$ and define

$$\begin{gathered} a = \epsilon/16, \qquad r_a = \frac{a\Delta_{\min}}{4\rho}, \qquad L_a = \frac{(1+a)^3}{(1-a)^2}T_\theta^\star, \\ m_a = \frac{1}{L_a\Delta_{\max}^2}, \qquad z_a = am_a, \qquad h(T) = \lceil\sqrt{T}\rceil, \qquad q(T) = q^-(T). \end{gathered} \tag{99}$$

Choose u_a so that for every $t \geq u_a$, $\eta_t \leq a\Delta_{\min}/(2\rho)$, and for every scheduled $u \geq u_a$, $\xi_u \leq a$. Such a deterministic u_a exists. Define $s_a(T) = \lceil h(T)/\kappa\rceil + 2$ and choose T_a large enough that, for every $T \geq T_a$,

$$\begin{gathered} h(T) \geq \max\{u_a, t_E, t_U, n_0\}, \qquad q(T) \geq \max\{h(T), T/2\}, \\ \frac{s_a(T) + n_0 + \sqrt{T}}{q(T)} \leq a. \end{gathered} \tag{100}$$

Existence follows from $s_a(T) = O(\sqrt{T})$ and $q(T)/T \to 1$. Thus these are deterministic, δ-independent thresholds for the fixed instance.

The two good events are

$$\begin{aligned} \mathcal{G}_T^{\mathrm{est}} &= \bigcap_{u=h(T)}^{T} \{\|\hat\theta_u - \theta\|_{\mathrm{loc}} \leq r_a\}, \\ \mathcal{G}_T^{\mathrm{sam}} &= \left\{\max_{(e,x)\in\mathcal{J}} |\bar\mu_{q(T),e,x} - \bar q_{q(T),e,x}| \leq z_a\right\}, \qquad \mathcal{G}_T = \mathcal{G}_T^{\mathrm{est}} \cap \mathcal{G}_T^{\mathrm{sam}}. \end{aligned} \tag{101}$$

Lemma 34 (Deterministic good-event implication)**.** *For $T \geq T_a$, after increasing T_a if necessary, $\mathcal{G}_T$ implies*

$$\frac{Z_{q(T)}}{q(T)} \geq (1-\epsilon)c_\theta. \tag{102}$$

Proof. (a) Target accuracy. On $\mathcal{G}_T^{\mathrm{est}}$, every scheduled $u \in [h(T), T]$ satisfies $(2\rho\|\hat\theta_u - \theta\|_{\mathrm{loc}} + \rho\eta_u)/\Delta_{\min} \leq a$. Lem. 24 ensures a unique quantized maximizer and gives

$$T_\theta(\mu_u) \leq (1+\xi_u)\left(\frac{1+a}{1-a}\right)^2 T_\theta^\star \leq \frac{(1+a)^3}{(1-a)^2}T_\theta^\star = L_a. \tag{103}$$

Thus every update in this interval succeeds, including updates whose targets are subsequently cached.

(b) Mass assigned to late targets. For any target-sampling round $s \geq s_a(T)$, the most recent scheduled update satisfies $v(s-1) \geq \kappa(s-1) \geq h(T)$. Hence the target used at s obeys (103). Put $t = q(T)$, $\mathcal{L} = \mathcal{I} \cap \{s_a(T), \ldots, t\}$, $\lambda = |\mathcal{L}|/t$, and $\nu = |\mathcal{L}|^{-1}\sum_{s\in\mathcal{L}} p_s$. Counting the omitted rounds and using (100),

$$\lambda \geq 1 - \frac{s_a(T) + R_t}{t} \geq 1 - \frac{s_a(T) + n_0 + \sqrt{T}}{q(T)} \geq 1 - a. \tag{104}$$

Convexity, the sublevel bound, and nonnegativity then imply

$$T_\theta(\nu) \leq L_a, \qquad \nu_{e,x} \geq m_a \quad ((e,x) \in \mathcal{J}_+), \qquad \bar q_{t,e,x} \geq \lambda\nu_{e,x}. \tag{105}$$

(c) Randomized tracking loss. On $\mathcal{G}_T^{\mathrm{sam}}$, each active factor marginal satisfies

$$\bar{\mu}_{t,e,x} \geq \bar{q}_{t,e,x} - z_a \geq \lambda \nu_{e,x} - a m_a \geq (\lambda - a)\nu_{e,x} \geq (1-2a)\nu_{e,x}. \tag{106}$$

The domination part of Lem. 23 gives the finite bound $T_\theta(\bar{\mu}_t) \leq L_a/(1-2a)$. No bound on the objective of early targets or forced point masses is needed.

(d) Empirical-parameter and stopping losses. Because $t = q(T) \geq h(T)$, the terminal local estimation error is at most r_a. In particular $2\rho r_a/\Delta_{\min} = a/2 \leq a$. Equations (53) and (106) yield

$$T_{\hat{\theta}_t}(\bar{\mu}_t) \leq \frac{T_\theta(\bar{\mu}_t)}{(1-a)^2} \leq \frac{L_a}{(1-2a)(1-a)^2}. \tag{107}$$

Both empirical maximizers equal $a^\star$ on this event. Also

$$\hat{\Delta}_{\min,t} \geq (1-a)\Delta_{\min}, \qquad \frac{2\rho\eta_t}{\hat{\Delta}_{\min,t}} \leq \frac{a}{1-a} \leq 2a < 1. \tag{108}$$

The check is therefore active, and (90) applies.

(e) Combined bound. Using $c_\theta = (2\sigma^2 T_\theta^\star)^{-1}$, the previous inequalities give

$$\begin{aligned} \frac{\underline{Z}_t/t}{c_\theta} &\geq (1-2a)^2 \frac{T_\theta^\star}{T_{\hat{\theta}_t}(\bar{\mu}_t)} \\ &\geq (1-2a)^3(1-a)^2 \frac{T_\theta^\star}{L_a} = \frac{(1-2a)^3(1-a)^4}{(1+a)^3} \\ &\geq (1-6a)(1-4a)(1-3a) \geq 1 - 13a \geq 1-\epsilon. \end{aligned} \tag{109}$$

Here $a \leq 1/32$. Bernoulli's inequality gives the first two factor bounds, convexity of $(1+a)^{-3}$ gives the third, and all lower factors are nonnegative. This proves (102). □

The complement of the good event is controlled without assuming independence between estimation and sampling errors. By (57) and the factor-coordinate version of Lem. 30,

$$\begin{aligned} \mathbb{P}_\theta(\mathcal{G}_T^c) &\leq \mathbb{P}_\theta((\mathcal{G}_T^{\mathrm{est}})^c) + \mathbb{P}_\theta((\mathcal{G}_T^{\mathrm{sam}})^c) \\ &\leq A_{r_a} T e^{-a_{r_a} c_0 T^{1/4}} + 2de^{-q(T)z_a^2/2} \\ &\leq A_{r_a} T e^{-a_{r_a} c_0 T^{1/4}} + 2de^{-Tz_a^2/4} =: b_\epsilon(T). \end{aligned} \tag{110}$$

Combining (102) and (110) gives, for every $T \geq T_a$,

$$\mathbb{P}_\theta\left(\frac{\underline{Z}_{q(T)}}{q(T)} < (1-\epsilon)c_\theta \right) \leq b_\epsilon(T), \qquad \sum_{T \geq T_a} b_\epsilon(T) < \infty. \tag{111}$$

To check summability explicitly, the first series is comparable to an integral and $\int_1^\infty x e^{-cx^{1/4}}\,dx = 4\int_1^\infty v^7 e^{-cv}\,dv < \infty$; the second is geometric.

E.5.4 Expected stopping time and completion of the theorem

Fix $\epsilon \in (0, 1/2)$ and write $c = (1-\epsilon)c_\theta$. For each fixed δ, $(a_\delta + g(q(T)))/q(T) \to 0$. Consequently there is a deterministic threshold after which $cq(T) > a_\delta + g(q(T))$. On $\mathcal{G}_T$, (102) forces stopping by $q(T) \leq T$. Equation (111) therefore proves $\mathbb{E}_\theta[\tau_\delta] < \infty$ for every fixed δ.

For the leading asymptotic constant, choose any $k > 1/c$ and set $n_\delta = \lceil kL_\delta \rceil$. For all sufficiently small δ, $n_\delta \geq T_a$ and $a_\delta > 3d$. By Lem. 28 and (94),

$$\frac{a_\delta + g(q(n_\delta))}{q(n_\delta)} = \frac{a_\delta + g(q(n_\delta))}{L_\delta} \frac{L_\delta}{n_\delta} \frac{n_\delta}{q(n_\delta)} \longrightarrow \frac{1}{k} < c. \tag{112}$$

The ratio $(a_\delta + g(t))/t$ is decreasing by (64), and $q(T)$ is nondecreasing. Therefore, for all these small δ,

$$a_\delta + g(q(T)) < cq(T) \quad \text{for every } T \geq n_\delta. \tag{113}$$

This is a uniform statement over the entire tail, not only at $T = n_\delta$. Since $\beta(q(T), \delta) \le a_\delta + g(q(T))$, equations (102) and (113) imply $\{\tau_\delta > T\} \subseteq \mathcal{G}_T^c$ for every $T \ge n_\delta$. The tail-sum identity now gives

$$\begin{aligned} \mathbb{E}_\theta[\tau_\delta] &= \sum_{T=0}^{\infty} \mathbb{P}_\theta(\tau_\delta > T) \\ &\le n_\delta + \sum_{T \ge n_\delta} b_\epsilon(T). \end{aligned} \tag{114}$$

with a tail tending to zero as $\delta \downarrow 0$. Hence

$$\limsup_{\delta \downarrow 0} \frac{\mathbb{E}_\theta[\tau_\delta]}{L_\delta} \le k \quad (k > 1/c), \qquad \limsup_{\delta \downarrow 0} \frac{\mathbb{E}_\theta[\tau_\delta]}{L_\delta} \le \frac{1}{(1-\epsilon)c_\theta}.$$

Let $\epsilon \downarrow 0$ to conclude

$$\limsup_{\delta \downarrow 0} \frac{\mathbb{E}_\theta[\tau_\delta]}{\log(1/\delta)} \le 2\sigma^2 T_\theta^\star. \tag{115}$$

The matching lower bound is the Gaussian MAMAB characteristic-time lower bound of Vannella et al. (2023a, Thm. 4.1), with the variance restored. Its normalization can be checked directly using Kaufmann et al. (2016, Lem. 1). For any uniformly δ-correct procedure with $m = \mathbb{E}_\theta[\tau] < \infty$ and any unique-optimum alternative λ, the stopped change-of-measure inequality gives

$$\sum_{e,x} \mathbb{E}_\theta[N_{\tau,e,x}] \frac{(\theta_{e,x} - \lambda_{e,x})^2}{2\sigma^2} \ge \mathrm{kl}(1-\delta, \delta), \qquad 0 < \delta < 1/2. \tag{116}$$

Here the event is recommendation of $a^\star$, and serialization gives the same stopped counts. The normalized expected counts $w_{e,x} = \mathbb{E}[N_{\tau,e,x}]/m$ are feasible factor marginals: expected global-action counts divided by m form a distribution whose projections are w. Let $\mu \in \mathcal{M}_O$ be the induced bag marginals. The weighted projection in Lem. 31, replacing counts by w, gives

$$\inf_{\lambda \in \mathrm{Alt}(\theta)} \sum_{e,x} \mu_{e,x} \frac{(\theta_{e,x} - \lambda_{e,x})^2}{2\sigma^2} = \min_{a \ne a^\star} \frac{\Delta_\theta(a)^2}{2\sigma^2 D_\mu^{a^\star}(a)} = \frac{1}{2\sigma^2 T_\theta(\mu)}. \tag{117}$$

The infimum is unchanged by closing the alternative half-spaces. To see this, take a point in their union and a maximizing action $a \ne a^\star$. Adding an arbitrarily small $\zeta > 0$ to each of its factor-configuration means makes a uniquely best: its score increases by $\rho\zeta$, while every distinct action receives strictly less. Distinct actions have distinct factor-configuration vectors, since uniqueness of any global optimum requires that every nonsingleton agent appears in a factor. This perturbation tends to zero with ζ. If a relevant weight is zero, the corresponding constraint can be met at zero cost, agreeing with the reciprocal conventions.

Combining (116)-(117) yields

$$m \ge 2\sigma^2 T_\theta(\mu)\, \mathrm{kl}(1-\delta, \delta) \ge 2\sigma^2 T_\theta^\star\, \mathrm{kl}(1-\delta, \delta).$$

Since $\mathrm{kl}(1-\delta, \delta)/L_\delta \to 1$, this lower bound and (115) prove

$$\lim_{\delta \downarrow 0} \frac{\mathbb{E}_\theta[\tau_\delta]}{\log(1/\delta)} = 2\sigma^2 T_\theta^\star. \tag{118}$$

Equations (92), (98), and (118) complete the proof of Thm. 10.

Probability and L^1 conclusions. Equation (98) implies convergence in probability. It also combines with (118) to give L^1 convergence. Indeed, for $X_\delta = \tau_\delta / L_\delta \ge 0$ and $c = 2\sigma^2 T_\theta^\star$,

$$\mathbb{E}_\theta |X_\delta - c| = \mathbb{E}_\theta X_\delta + c - 2\mathbb{E}_\theta \min(X_\delta, c) \longrightarrow c + c - 2c = 0,$$

where the last expectation converges by bounded convergence. All limits are pointwise in the fixed instance. The deterministic threshold T_a may depend strongly on $\Delta_{\min}$.

E.6 PROOF OF THE COMPUTATIONAL-COST STATEMENT

The following bounds include construction of local potentials and all intermediate tables. Let $B = |\mathcal{C}_O| \le N$ for the maximal-bag construction. The computational model counts arithmetic operations and array accesses; SOCP dimensions are not a bound on numerical solver runtime.

Lemma 35 (Uniqueness checks by top-two elimination)**.** *Given bag-local potentials, the best and second-best distinct assignments, including a possible tie in their scores, can be computed in $O(NK^{w_O+1})$ work. Constructing the bag potentials adds $O(\rho K^{w_O+1})$ work. Separator messages store two scores and their witnesses per assignment, using $O(NK^{w_O})$ message entries.*

Proof. For each subtree and fixed separator assignment, replace a scalar max-sum entry by its two highest-scoring *distinct completions*. Break score ties by a fixed candidate-generation order. Conditional on a fixed bag assignment, two child subproblems have disjoint residual variables. If their ordered completion scores are $x_1 \ge x_2 \ge \cdots$ and $y_1 \ge y_2 \ge \cdots$, their two best combined witnesses belong to

$$\{(1,1),(1,2),(2,1),(2,2)\}, \qquad \text{with scores} \quad \{x_i + y_j : 1 \le i,j \le 2\}. \tag{119}$$

Indeed, if a candidate uses rank $i \ge 3$ in the first component, replacing that component by ranks one and two gives two distinct witnesses with no smaller score. The argument for the second component is identical. This remains true at ties and when one component has only one completion.

Merge children sequentially, retaining two witnesses after each constant-sized merge. Add the bag score, and then maximize over bag assignments compatible with each parent separator. Distinct residual bag assignments give distinct completions; for a fixed residual assignment, the child witnesses are distinct by induction. Hence retaining the top two through this projection is exact. The leaf case starts the induction. At the root the two best distinct global assignments have equal scores precisely when the maximum is not unique.

The scopes are those of ordinary JT max-sum (Wainwright and Jordan, 2008, Sec. 2.5.2). At most K^{w_O+1} bag assignments are inspected at each bag. There are $B-1$ child incidences, and each merge has constant overhead. Thus

$$\text{Work} \le O\left(\sum_C (1 + |\operatorname{ch}(C)|)|\mathcal{A}_C|\right) = O(NK^{w_O+1}).$$

Summing the tables of hosted factors costs $O(\sum_C |\mathcal{E}_C||\mathcal{A}_C|) \le O(\rho K^{w_O+1})$. Adjacent distinct inclusion-maximal bags have separator size at most w_O, yielding $O(NK^{w_O})$ score entries. Constant-rank backpointers recover the two witnesses in a further tree traversal; full actions need not be copied into each table entry. □

Lemma 36 (Lattice width of all intermediate gap tables)**.** *For local means $\vartheta_{e,x} \in \eta\mathbb{Z}$ and any reference action b, let $B_e^\vartheta = \max_x \vartheta_{e,x} - \min_x \vartheta_{e,x}$. Every message, sequential child-combination table, and root table of the gap recursion has, at any fixed non-gap assignment, at most*

$$1 + \left\lfloor \frac{1}{\eta} \sum_e B_e^\vartheta \right\rfloor \tag{120}$$

attainable gap values. The same bound suffices for a dense interval-indexed implementation.

Proof. At a fixed stage of a sequential child combination, a fixed subset $F \subseteq [\rho]$ of factors has been included, with each factor counted once. Its partial gap is $s = \sum_{e\in F}(\vartheta_{e,b_e} - \vartheta_{e,a_e})$. For $l_e = \vartheta_{e,b_e} - \max_x \vartheta_{e,x}$ and $u_e = \vartheta_{e,b_e} - \min_x \vartheta_{e,x}$,

$$s \in \eta\mathbb{Z} \cap \left[\sum_{e\in F} l_e,\ \sum_{e\in F} u_e\right], \qquad \sum_{e\in F}(u_e - l_e) = \sum_{e\in F} B_e^\vartheta \le \sum_e B_e^\vartheta. \tag{121}$$

Fixing a bag or separator assignment can only remove feasible completions. An interval of width W contains at most $1 + \lfloor W/\eta \rfloor$ points of $\eta\mathbb{Z}$. This proves the claim for every intermediate stage, not only completed separator messages. A dense array indexed by the whole interval in (121), with $-\infty$ at unreachable entries, uses the same bound. Negative partial sums require an index offset, not extra states or a positivity restriction. □

Lemma 37 (Uniform empirical-range envelope)**.** *Let* $\bar{X}_{e,x,n} = n^{-1}\sum_{j=1}^{n} X_{e,x,j}$ *and define*

$$R = \max_{(e,x)\in\mathcal{J}} \sup_{n\geq 1} |\bar{X}_{e,x,n} - \theta_{e,x}|.$$

Then $\mathbb{E}_\theta R^p < \infty$ *for every fixed* $p > 0$*. For every* $t \geq n_0$*,*

$$B_{t,e}^{\eta_t} := \max_x \hat{\theta}_{t,e,x}^{\eta_t} - \min_x \hat{\theta}_{t,e,x}^{\eta_t} \leq B_e + 2R + \eta_t, \qquad B_e = \max_x \theta_{e,x} - \min_x \theta_{e,x}. \tag{122}$$

Proof. Applying (59) to deterministic stream prefixes and summing over all n gives, for $z > 0$,

$$\mathbb{P}_\theta(R > z) \leq \min\left\{1,\ 2d\sum_{n\geq 1} e^{-nz^2/(2\sigma^2)}\right\} = \min\left\{1,\ \frac{2de^{-z^2/(2\sigma^2)}}{1-e^{-z^2/(2\sigma^2)}}\right\}. \tag{123}$$

For any fixed $z_0 > 0$, the denominator on $z \geq z_0$ is bounded below. Hence, with $C_{z_0} = 2d/(1 - e^{-z_0^2/(2\sigma^2)})$,

$$\mathbb{E}_\theta R^p = \int_0^\infty pz^{p-1}\mathbb{P}_\theta(R > z)\,dz \leq z_0^p + C_{z_0}\int_{z_0}^\infty pz^{p-1}e^{-z^2/(2\sigma^2)}\,dz < \infty.$$

Every adaptive empirical mean is one of the stream averages entering R. Thus $\max_x \hat{\theta}_{t,e,x} - \min_x \hat{\theta}_{t,e,x} \leq B_e + 2R$. Rounding changes either endpoint by at most $\eta_t/2$, proving (122) uniformly over all times. □

Proof of the computational-cost statement. Let H_t be the largest attainable gap count over all intermediate tables, and let $\bar{H}_t$ be the sufficient dense-grid width of Lem. 36. Lemmas 36 and 37 yield

$$\begin{aligned} H_t \leq \bar{H}_t &\leq 1 + \frac{1}{\eta_t}\sum_e B_{t,e}^{\eta_t} \\ &\leq 1 + \rho + \frac{\log(t+2)}{c_\eta}\left(\sum_e B_e + 2\rho R\right). \end{aligned} \tag{124}$$

The random coefficient is finite almost surely and has moments of every order. Consequently, for each fixed instance, $\bar{H}_t = O(\log(t+2))$ a.s. and $\mathbb{E}_\theta[\bar{H}_t^p] = O(\log^p(t+2))$ for every fixed $p > 0$.

For completeness, implement (87) without enumerating a Cartesian product of all child gap sets. Order the children of C as $D_1, \ldots, D_m$. For a fixed $y \in \mathcal{A}_C$, initialize $F_{C,0}(y,s) = c_{C,y}$ if $s = g_{C,y}$, and $-\infty$ otherwise. Then use the sequential max-convolution

$$F_{C,j}(y,s) = \max_{s'+s''=s}\left\{F_{C,j-1}(y,s') + M_{D_j}(y_{C\cap D_j}, s'')\right\}, \qquad M_C(z,s) = \max_{y:y_{S_C}=z} F_{C,m}(y,s). \tag{125}$$

Every pair of reachable indices determines one sum index, so a merge costs at most $O(\bar{H}_t^2)$ per bag assignment. There are $B - 1 \leq N - 1$ child merges. Constructing the bag-local gaps and costs costs $O(\rho K^{w_O+1})$, and projection costs $O(NK^{w_O+1}\bar{H}_t)$. Therefore

$$\begin{aligned} \mathrm{Work}_{\mathrm{stop}}(t) &= O\big((N\bar{H}_t^2 + \rho)K^{w_O+1}\big), \\ \mathrm{Storage}_{\mathrm{stop}}(t) &= O\big((N\bar{H}_t + \rho)K^{w_O+1}\big). \end{aligned} \tag{126}$$

The ρ term accounts for forming bag potentials and may be omitted only if they are already supplied. The input statistic updates cost $O(\rho)$ per global round. Computing empirical and quantized uniqueness by Lem. 35 is covered by (126).

At a target update, substituting $H_t \leq \bar{H}_t$ into the static formulation of Thm. 5 gives

$$\#\mathrm{variables} = O\big((N\bar{H}_t + \rho)K^{w_O+1}\big), \qquad \#\mathrm{constraints} = O\big((N\bar{H}_t^2 + \rho)K^{w_O+1}\big). \tag{127}$$

These are formulation sizes. The statistical theorem assumes feasible outputs satisfying (49); it does not assert that an arbitrary floating-point solver returns exact optima in the time in (127).

For the specific choices $\mathcal{U} = \{n_0\} \cup \{2^k : 2^k \geq n_0\}$ and $\mathcal{Q} = \{m^2 : m^2 \geq n_0\}$,

$$|\mathcal{U} \cap [1,T]| \leq 2 + \lfloor \log_2 T \rfloor, \qquad |\mathcal{Q} \cap [1,T]| \leq \lfloor \sqrt{T} \rfloor.$$

Summing (126) at $t = m^2$ yields

$$\sum_{m^2 \le T} \mathrm{Work}_{\mathrm{stop}}(m^2) = O\left(K^{w_O+1} \sum_{m \le \sqrt{T}} [N \bar{H}^2_{m^2} + \rho] \right) = O\Big(K^{w_O+1} [N\sqrt{T} \log^2(T+2) + \rho\sqrt{T}] \Big). \tag{128}$$

The a.s. implicit coefficient is bounded by a constant times $1 + R^2$, so the same order also holds in expectation. These aggregate stopping costs are $o((N+\rho)K^{w_O+1}T)$ for a fixed instance. No aggregate numerical runtime for target optimization is inferred from its logarithmic number of calls.

Given a cached target, (69) requires one root categorical draw and one conditional draw per nonroot bag. Conditional tables can be formed by marginalization in $O(\sum_C |\mathcal{A}_C|)$ operations. Thus sampling uses B categorical primitives per target-sampling round. With cumulative-probability tables, binary search gives $O(\sum_C \log |\mathcal{A}_{C \setminus S_C}|)$ comparisons for nonroot draws, plus the root draw; the number of categorical primitives should not be confused with their implementation cost. Full empirical bag counters, if retained, require one coordinate increment per bag per round. Factor counts and reward sums alone suffice for evaluating the likelihood and its factor projections. □

Scope of the logarithmic-resolution bound. The statistical limits fix θ before sending $\delta \downarrow 0$. They are not uniform over $\Delta_{\min}$. In particular,

$$\rho\eta_t \le r \quad \Longleftrightarrow \quad t + 2 \ge \exp(\rho c_\eta / r). \tag{129}$$

For one factor with two means $(\Delta, 0)$, rounding the true table gives a tie whenever $\eta_t > 2\Delta$, equivalently $t + 2 < \exp(c_\eta/(2\Delta))$. This is a deterministic resolution example, not a pathwise stopping-time lower bound for noisy estimates. It shows why logarithmic state growth does not imply a gap-uniform transient guarantee. Other deterministic vanishing grids satisfy the consistency arguments above, but have different state-growth bounds.

E.7 Proof of the coordinator-assisted distributed-realization statement

Assume synchronized global rounds, reliable communication along the JT, and the same measurable oracle selection and numerical tolerances as in the centralized procedure. Each factor has one designated host. The allocation SOCP is assembled and solved by the coordinator; only the tree passes and action generation are distributed. We prove equality of the complete experiment laws, not merely equality of optimum objective values.

Proof of distributed equivalence. Couple the procedures using the same factor reward streams and, for each target-sampling round, one common seed for each root/child categorical draw. Use the same initialization, exploration order, schedules, and tie rules. Let H_t^{cen} and H_t^{dist} be their histories through round t. We prove by induction that

$$H_t^{\mathrm{cen}} = H_t^{\mathrm{dist}}, \qquad \mu_t^{\mathrm{cen}} = \mu_t^{\mathrm{dist}}, \qquad Z_t^{\mathrm{cen}} = Z_t^{\mathrm{dist}}, \qquad \underline{Z}_t^{\mathrm{cen}} = \underline{Z}_t^{\mathrm{dist}} \tag{130}$$

whenever the corresponding target or statistic is computed. Initialization gives equality at $t = n_0$.

Assume the invariant through t. The host of (e, x) has the same revealed rewards and counts, hence

$$N_{t,e,x}^{\mathrm{cen}} = N_{t,e,x}^{\mathrm{dist}}, \quad \hat{\theta}_{t,e,x}^{\mathrm{cen}} = \hat{\theta}_{t,e,x}^{\mathrm{dist}}, \quad \hat{\theta}_{t,C,y}^{\mathrm{cen},\eta_t} = \hat{\theta}_{t,C,y}^{\mathrm{dist},\eta_t}.$$

At a target update, these values produce identical local blocks of the same SOCP. The shared oracle-selection rule returns the same feasible marginal collection; at a tie both procedures retain the same previous target. At a stopping check, induction from leaves to root in (87) shows equality of every separator-gap message, hence equality of the root profile and of $\underline{Z}_t$. Uniqueness checks use the same local potentials and tie rule. Thus both procedures make the same stopping decision.

If a new action is needed, the initialization or forced branch is deterministic and identical. In the target branch, the coupled root seed draws the same root assignment. Conditional on equal parent assignments, the child kernels in (69) agree. The coupled child seed therefore gives the same residual assignment. Induction down the tree gives $A_{t+1}^{\mathrm{cen}} = A_{t+1}^{\mathrm{dist}}$. The identical action reveals identical next unused reward-stream entries. This proves the invariant at $t + 1$.

It follows pathwise under the coupling that, for every confidence level,

$$\big((A_t, Y_t)_{t\ge 1}, \tau_\delta, \hat{a}_{\tau_\delta}\big)^{\mathrm{cen}} = \big((A_t, Y_t)_{t\ge 1}, \tau_\delta, \hat{a}_{\tau_\delta}\big)^{\mathrm{dist}}.$$

Consequently their laws and statistical guarantees coincide.

The communication sizes also follow directly from the state domains. A top-two message has at most $2|\mathcal{A}_{S_C}| \le 2K^{w_O}$ records. A stopping message has at most $|\mathcal{A}_{S_C}|\bar{H}_t \le K^{w_O}\bar{H}_t$ scalar entries. Thus complete upward passes communicate $O(NK^{w_O})$ and $O(NK^{w_O}\bar{H}_t)$ scalar entries, respectively. A downward sampling traversal communicates only the realized separator assignments, requiring at most

$$\sum_{C\neq R} |S_C|\lceil \log_2 K\rceil \le (N-1)w_O\lceil \log_2 K\rceil$$

bits, up to message headers, when agents use common action encodings. Communication with the centralized conic solver is separate from these tree-pass counts. □

F ADDITIONAL EXPERIMENTAL RESULTS

This appendix provides additional experimental results and details supporting the results in Sec. 7. We first describe the controlled instances and the allocation geometry used throughout the online experiments. We then report additional characteristic-time comparisons across random instances and graph topologies, finite-sample results across confidence levels, stopping-rule calibration, mean-field diagnostics, quantization effects, and computational scaling. We finally consider random graph families beyond the canonical topologies used in the main experiments. Unless stated otherwise, factor rewards are Gaussian with standard deviation $\sigma = 0.5$. All experiments use deterministic seeds, and individual runs are stored before aggregation. The main online benchmark consists of five $N = 5$ instances rescaled to the same minimum gap $\Delta_{\min} = 0.35$.

F.1 CONTROLLED INSTANCES AND ALLOCATION GEOMETRY

The instances are designed to vary the aspects of the problem that determine the statistical and computational behavior of BAI while keeping the minimum reward gap fixed. It contains a correlation-sensitive pairwise chain, a heterogeneous-gap pairwise chain, a binary cycle, a model with overlapping third-order factors, and a ternary-action chain. Together these instances vary dependence structure, gap geometry, factor order, induced width, and local action-space size.

For the online comparisons, we consider JT-TaS, COMBGAME-OFW (Jourdan et al., 2021), common-gap JT, and Uniform global sampling under the same experimental stopping calibration. Oracle-JT uses the true target allocation and isolates the cost of online target estimation. Oracle-MF is given the true factor means and therefore isolates the statistical effect of restricting the allocation to an independent mean-field family. At $\delta \in \{0.10, 0.05\}$, we use common random seeds with 20 repetitions per method and instance. Fig. 3 summarizes the controlled families used in the main experiments presented in Sec. 7.

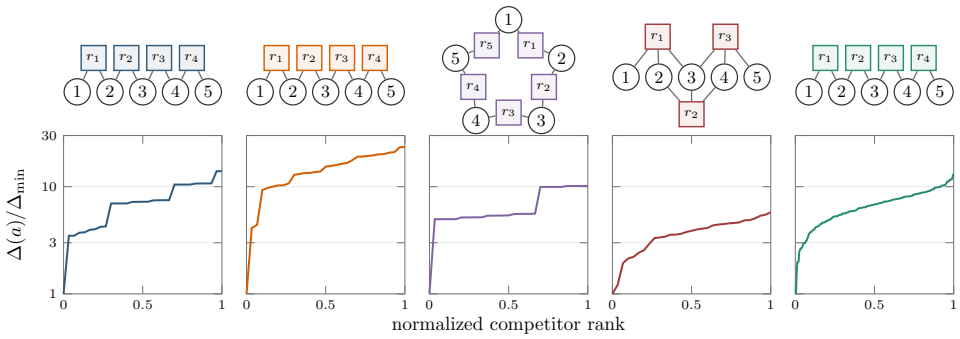

Figure 3: **Controlled benchmark instances.** The main instances used in the online experiments (Sec. 7 and their normalized competitor ranks.

F.2 CHARACTERISTIC-TIME ACCURACY ACROSS GRAPH TOPOLOGIES

We next test whether the characteristic-time accuracy observed on the controlled instances persists across different graph structures and reward realizations. We consider seven graph families: chain, balanced tree, star, ring, 2×4 ladder, 3×3 grid, and overlapping triplet chain, with $N \in \{7, 8, 9\}$. For each family, we generate 10 independent instances with a unique optimal action and rescale the rewards to $\Delta_{\min} = 0.35$. This gives 70 instances in total, all small enough for exhaustive action-indexed optimization to provide a numerical reference for $T^\star_\theta$.

Figure 4 reports the target value obtained by JT, MF, and common-gap JT relative to the exhaustive characteristic time. Each box summarizes the 10 instances of one graph family; the central line denotes the median, the box the interquartile range, and the whiskers extend to 1.5 times the interquartile range, with remaining points shown as outliers. The vertical axis is logarithmic.

JT closely matches the exhaustive solution across every topology and reward realization: over all 70 instances,

$$1.00006 \leq \frac{T_{\mathrm{JT}}}{T_\theta^\star} \leq 1.00157.$$

Thus, the agreement observed on the controlled instances is not specific to their particular topology or gap geometry. In contrast, restricting the allocation to the independent MF family or replacing the action-dependent gaps by a common gap can substantially increase the characteristic time. Across these instances, the largest observed ratios are $21.44\times$ for MF and $17.82\times$ for common-gap JT.

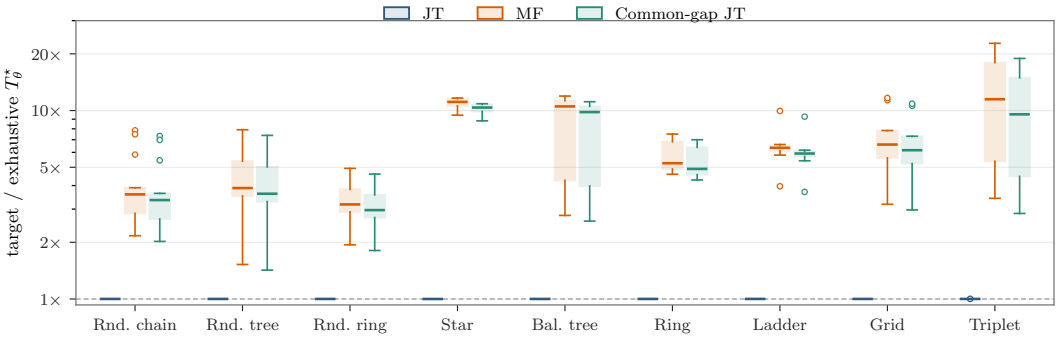

Figure 4: **Characteristic-time accuracy across graph topologies.** Target value relative to exhaustive $T_\theta^\star$ across the seven graph families.