Leveraging Similarities in Multi-Armed Bandits

Khaled Eldowa KHALED.ELDOWA@INRIA.FR
Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 38000 Grenoble, France

Thibaud Rahier T.RAHIER@CRITEO.COM
Criteo AI Lab, Paris, France

Augustin Cablant AUGUSTIN.CABLANT@INRIA.FR
Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 38000 Grenoble, France

Panayotis Mertikopoulos PANAYOTIS.MERTIKOPOULOS@IMAG.FR
Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LIG, 38000 Grenoble, France

Pierre Gaillard PIERRE.GAILLARD@INRIA.FR
Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 38000 Grenoble, France

Abstract

In many online learning and bandit problems, the actions we consider possess inherent similarities–for instance because they share latent traits, tags, or hierarchical structure. We study online learning with a similarity-structured action set, encoded by a rooted tree whose leaves are the actions and whose levels quantify how closely two actions are related. The loss sequence is assumed tree-compatible: losses of similar actions are constrained to be close. We establish an impossibility result showing that usual one-point bandit feedback cannot, in general, leverage range or tree-induced similarity, even under very strong similarity constraints. We then provide a unified set of algorithms which adapt to a wide range of richer feedback models, from semi-bandit feedback down to multi-point bandit protocols, including the minimal two-point feedback setting. We show these algorithms exhibit best-of-both-worlds guarantees and provably exploit action similarities by replacing the number of actions K by a similarity-aware effective number of actions K_{eff} in the regret bounds. As an application, we show that under two-point feedback, it is possible to achieve $\sqrt{T}$ regret in Lipschitz bandits when $d \leq 2$.

Keywords: Online Learning, Structured Bandits, Multi-Point Feedback, Best-of-Both-Worlds.

1. Introduction

Consider the following recommendation problem: An ad placement platform has a catalogue of products—let's say cars for concreteness—and is called to make product recommendations to a stream of users connecting to the system. This catalogue may have an innate structure: for example, cars may be classified according to brand (Ford, Chevrolet, Toyota, Peugeot,...), type (SUVs, compacts, minivans,...), propulsion (gas, hybrid, or electric), and/or any other relevant trait. In turn, this categorization across different traits (brand, type, propulsion, ...) may have a commensurate effect on the value of each product: thus, a user looking to buy a large family car might have a relatively high value for all mid-size cars in the catalogue, a somewhat lower value for SUVs, and a far lower value for compacts; and, depending on the user's criteria on the catalogue's other traits, this could further constrain the value range of each combination of traits.

Seen as a repeated decision problem, the learner here seeks to minimize their regret in the presence of a large number of distinct alternatives, some of which may be more "similar" than others. These similarities are defined by means of a family of traits which group together alternatives

© K. Eldowa, T. Rahier, A. Cablant, P. Mertikopoulos & P. Gaillard.

that share a given attribute (or more). More precisely, each trait gives rise to a partition of the set of possible choices, and each partition (or combination thereof) might greatly constrain the range of rewards that can be observed within: for instance, the value of two hybrid SUVs of the same make and model but different year would be highly correlated; by contrast, the rewards associated to a hybrid compact and a hybrid SUV of different make could diverge considerably. A natural goal here is to determine the degree to which such innate similarities within the set of alternatives may lead to improved regret guarantees, and how to achieve them.

A natural way to operationalize such a structure is to encode it as a hierarchy, i.e., a tree-structured notion of similarity over the alternatives. This presupposes fixing a priority ranking over the traits; e.g., 1) brand, 2) type, 3) propulsion, etc. Then, the traits from the first up to, say, the n-th collectively partition the alternatives, with each set of this partition corresponding to a possible combination of domain values that this group of traits can take. This partition in turn refines the one induced by the traits from 1 up to $n-1$. Returning to our example, the combination of "brand" and "type" partitions the set of alternatives (here, cars) into "Ford sedans", "Ford SUVs", "Toyota sedans", etc; which is a refinement of the partition induced by "brand" alone: "Ford", "Toyota", etc. Ultimately, one obtains a nested sequence of partitions of the set of alternatives, which can be represented as a tree whose root is the entire set of alternatives and whose leaves correspond to individual alternatives. See Appendix A for a more formal exposition of this process.

Crucially, the difference in outcome between two alternatives (in terms of gains or losses) can now be simply bounded as a function of the height of their lowest common ancestor in the tree; the lower it is, the more traits they agree on. Motivated by this example, and taking a more abstract point of view, we study in this work a multi-armed bandit problem where the actions are the leaves of a given tree, and where the losses of any two actions can diverge by no more than a given function of the height of their most recent common ancestor. The main question we seek to answer is: *when* can these similarity constraints translate into strictly improved regret guarantees, and *what feedback is needed* to actually exploit them?

The first point we establish is that, perhaps surprisingly, exploiting this structure is not possible in general under standard 1-point bandit feedback. This is in contrast to parametric bandit models like linear bandits (Bubeck et al., 2012), where one can improve upon the minimax regret rate of unstructured bandits. We prove this impossibility result by building upon the fact that one cannot exploit the effective range of the losses in (adversarial) multi-armed bandits (Gerchinovitz and Lattimore, 2016); that is, the maximum possible difference between the losses of two actions.

It is no surprise then that the closely related *nested bandit* model of Martin et al. (2022) assumes access to more informative feedback. In that work, they also consider a tree-based representation with leaves as actions. They assume that the loss of an action admits a nested decomposition along the path to the root of the tree, with the expectation that the intrinsic losses are smaller for nodes at lower levels as they represent less relevant characteristics of the action. They adopt a semi-bandit feedback model: upon choosing an action, all node-level losses along the path to the root are revealed. The regret bounds they achieve depend on an *effective number of actions* K_{eff} (which is a function of the importance and size of intermediate node sets) rather than on the total number of actions K. However, in many applications, this semi-bandit feedback is difficult to justify; in the car-catalogue example, one typically observes only an overall satisfaction signal for the recommended car, not separate losses attributable to brand, model, color, etc. As a second limitation, they only focus on adversarial losses and do not provide guarantees in the stochastic setting.

Here, we address both limitations. We develop a best-of-both-worlds approach that yields guarantees in both stochastic and adversarial regimes depending on notions of effective number of actions, and we substantially weaken the feedback requirements by moving to multi-point protocols–culminating in a minimal two-point feedback model in which only the losses of two carefully coupled actions are observed at each round. Compared to semi-bandit feedback, it is often more realistic to obtain such multi-point feedback–for instance from several users making different choices. Likewise, in routing and path-finding applications, one may send several drivers along different routes and observe their corresponding total travel times. Importantly, our feedback model does not provide free exploration; the losses of all chosen actions are accounted for in the regret.

In the final part of this work, we show that, beyond the tree-structured model, multi-point feedback can also be advantageous in the more familiar Lipschitz bandits problem (Kleinberg et al., 2008). In particular, under two-point feedback, we obtain new regret guarantees that, in lower dimensions, are significantly better than what can be achieved under one-point feedback. These new guarantees are obtained by applying our techniques to a hierarchical discretization of the action space.

We now summarize our contributions.

1.1. Contributions

Our contributions can be summarized as showing that action similarities can be leveraged, under a relatively weak feedback model, to obtain strictly sharper best-of-both-worlds bounds compared to standard bandit guarantees.

(1) A lower bound for one-point bandit feedback. We prove that one-point (bandit) feedback cannot, in general, exploit tree-induced similarity constraints: for any such similarity structure, there always exist loss sequences that force $\Omega(\sqrt{KT})$ regret. This motivates considering feedback models that reveal information from at least two actions per round.

(2) Best-of-both-worlds guarantees under tree-induced similarities. We adopt a follow-the-regularized-leader algorithmic template with a tree-aware regularizer that achieves logarithmic-type guarantees in stochastic regimes and $\sqrt{T}$-type guarantees in adversarial regimes. In both cases, the similarity structure yields improvements over structure-agnostic baselines by replacing the ambient dependence on K with an *effective* number of actions K_{eff} (and, in the stochastic case, an effective gap-dependent complexity), which can be dramatically smaller than K when losses vary smoothly across the action set.

(3) Multi-point feedback relaxations, up to a minimal two-point protocol. We consider multi-point bandit protocols ($m \geq 2$) ranging from an $(L+1)$-point scheme (where L is the depth of the tree) aligned with the hierarchy down to a *minimal* two-point feedback model in which only the losses of two carefully coupled actions are observed per round. We show that we can achieve guarantees comparable to those that can be obtained under the semi-bandit model of (Martin et al., 2022).

(4) Application to Lipschitz bandits. As an application, we show that under two-point feedback, our techniques can yield interesting results in the Lipschitz bandits problem, where we show that it is possible to achieve $\sqrt{T}$ regret when $d \leq 2$.

1.2. Related Work

Multi-armed bandits are a standard framework for sequential learning under uncertainty and have been extensively studied. We refer to Lattimore and Szepesvári (2020) for a comprehensive overview and focus here on the most relevant works for our setting.

Exploiting structure in multi-armed bandits. Classical bandit models typically assume a finite set of independent arms. However, many modern applications involve structured action sets in which arms are related through geometry, correlations, or combinatorial constraints. This has motivated a substantial body of work on bandits with structured action spaces. Structured bandits extend the basic multi-armed bandit framework to large or infinite arm sets by imposing known relationships on the reward function across arms, which can be exploited to improve learning performance. Prominent examples include linear bandits (Abbasi-Yadkori et al., 2011; Agrawal and Goyal, 2013; Ito et al., 2020) and combinatorial bandits (Wen et al., 2015; Combes et al., 2015). A unifying treatment of several such structures in the stochastic setting was proposed by Combes et al. (2017). Another important class is given by Lipschitz or Hölder bandits, where arms lie in a metric (possibly continuous) space and regret bounds depend on the covering or packing dimension of this space (Kleinberg et al., 2008; Slivkins, 2011; Kleinberg et al., 2019; Podimata and Slivkins, 2021).

In contrast to these parametric or metric-based models, we consider a setting where the similarity structure of the action space is encoded by a tree. This representation is motivated both by Lipschitz bandits (see Sec. 7) and by applications involving catalogues of objects that share hierarchical attributes (see Sec. 2). Such implicit tree structures have previously been studied by Slivkins (2011), who refer to them as taxonomies of arms, in the stochastic setting, and by Martin et al. (2022) for adversarial bandits with semi-bandit feedback. Another relevant line of work is hierarchical stochastic bandits (e.g., Hong et al. 2022a,b; Aouali et al. 2023; Nguyen et al. 2025), where correlations between rewards are imposed via hierarchical Bayesian models. Hong et al. (2022a), in particular, consider a tree-structured action set where nearby actions can have strongly correlated rewards. When the structure is favorable, they obtain better regret bounds compared to structure-agnostic algorithms. However, they only consider a Bayesian setting and do not provide $\ln T$ bounds.

Finally, structural assumptions cannot always be fully exploited under standard bandit feedback. In particular, Gerchinovitz and Lattimore (2016) show that adaptation to the intrinsic loss range is impossible. We extend this limitation to our tree-structured setting in Section 3, motivating our multi-point feedback model.

Multi-point bandit feedback. Several existing works have shown that improved regret guarantees can be obtained when the learner observes the loss at additional points at each round. Such improvements date back to Agarwal et al. (2010), which improves the regret from $O(T^{3/4})$ to $O(\sqrt{T})$ compared to the original one-point feedback gradient descent method of (Flaxman et al., 2004; Kleinberg, 2004) for bandit convex optimization. Subsequent works (Duchi et al., 2015; Shamir, 2017; Lu et al., 2024) similarly demonstrate the benefits of two-point feedback for reducing the variance of gradient estimates or compete with non-stationary bandit environments. However, these approaches rely on gradient estimation techniques and assume convex losses, which is not the case in our setting. In the multi-armed bandit literature, observing an additional point has been studied by Sani et al. (2014) to adapt to "easy data" and by Thune and Seldin (2018) to adapt to the effective range of the losses, thereby circumventing the impossibility result of Gerchinovitz and Lattimore

(2016) for one-point feedback. Additionally, Degenne et al. (2018) study a stochastic bandit problem with occasional (free) additional observations, which allows obtaining constant regret bounds. In our case, we show that two-point feedback allows us to exploit the tree structure of the action set. However, unlike in (Sani et al., 2014), (Thune and Seldin, 2018), and (Degenne et al., 2018), this additional observation does not constitute free exploration; rather, it is explicitly accounted for in our regret definition, which is better motivated by practical considerations.

2. Preliminaries

A Tree-Based Action Set We consider an action set $\mathcal{A} \coloneqq \{1, \dots, K\}$. We will represent $\mathcal{A}$ as the set of leaves of a given, fixed rooted tree $\mathcal{T}$ of depth L. We summarize now some tree-related notation. For each level $j \in \{0, 1, \dots, L\}$, let V_j be the set of nodes at depth j, with $V_0 = \{v_0\}$ (the root) and $V_L = \mathcal{A}$. We define $\mathcal{V} \coloneqq \cup_{j=1}^{L} V_j$ and $\overline{\mathcal{V}} \coloneqq \cup_{j=0}^{L} V_j$. We use $\triangleright$ to denote the parenthood relation: $v \triangleright w$ means that w is a child of v. This can only happen if $v \in V_{j-1}$ and $w \in V_j$ for some $j \in [L]$. We assume that every node has at least one child until level L: for any $j \in \{0, \dots, L-1\}$ and $v \in V_j$, there exists at least one $w \in V_{j+1}$ such that $v \triangleright w$. In particular, $v_0 \triangleright w$ holds for every $w \in V_1$.

Let $\succ$ denote the reflexive and transitive closure of $\triangleright$ (ancestry relation), so that $v \succ w$ means that v is an ancestor of w (and every node is its own ancestor). Additionally, the function $\mathrm{lvl}\colon \overline{\mathcal{V}} \to [L] \cup \{0\}$ yields for any $v \in \overline{\mathcal{V}}$ the unique level j such that $v \in V_j$. Moreover, for every $j \in [L] \cup \{0\}$, the function $\mathrm{anc}_j\colon \bigcup_{l \geq j} \mathcal{V}_l \to \mathcal{V}_j$ maps every node at level j or lower to its unique ancestor at level j. The function $\mathrm{prt}\colon \mathcal{V} \to \overline{\mathcal{V}}$ maps a node v to its unique parent $\mathrm{prt}(v)$. For each $j \in [L] \cup \{0\}$, we define the equivalence relation $\sim_j$ on $\bigcup_{l \geq j} V_l$ by

$$v \sim_j w \quad \iff \quad \mathrm{anc}_j(v) = \mathrm{anc}_j(w).$$

When restricted to leaves, $a \sim_j b$ means that a and b lie in the same subtree rooted at $\mathrm{anc}_j(a)$. We encode the hierarchical proximity between actions through their divergence level, using a *similarity* function $s : \mathcal{A} \times \mathcal{A} \to \{1, \dots, L+1\}$ defined as

$$s(a, b) \coloneqq 1 + \max\{j \in \{0, \dots, L\} \mid a \sim_j b\}, \qquad a, b \in \mathcal{A}.$$

Equivalently, for two actions $a, b \in \mathcal{A}$, $s(a, b)$ is the first level at which a and b diverge ($s(a, a) = L + 1$), so the larger $s(a, b)$, the more similar a and b are. We identify functions $x : \mathcal{A} \to \mathbb{R}$ with vectors in $\mathbb{R}^K$ through an arbitrary (fixed) bijection between $\mathcal{A}$ and $[K]$. The same applies to probability distributions over $\mathcal{A}$, which we view as vectors in Δ_K, the probability simplex in $\mathbb{R}^K$. For any vector $x \in \mathbb{R}^K$ and any node $v \in \overline{\mathcal{V}}$, we define the subtree aggregate

$$x[v] \coloneqq \sum_{a \in \mathcal{A}\colon\, v \succ a} x(a),$$

i.e., the mass (or value) of x restricted to the subtree rooted at v.

Tree-compatible loss model This work is framed under a sequential online bandit problem (described in more detail later) in which the learner faces a sequence of loss functions $(y_t)_{t \geq 1}$ with $y_t : \mathcal{A} \to \mathbb{R}$ generated by the environment. We will assume that the loss functions respect the similarity between actions in the following sense: we assume there exist parameters $(\sigma_j)_{j \in [L+1]}$ (with $\sigma_{L+1} = 0$) such that for all rounds t and all $a, b \in \mathcal{A}$,

$$|y_t(a) - y_t(b)| \leq \sigma_{s(a,b)}. \tag{1}$$

That is, pairs sharing a deeper common ancestor have more similar losses. We posit, naturally, that $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_L$. Additionally, for the similarity structure to provide an advantage to the learner, these values need to decrease sufficiently fast compared to the growth of the number of nodes in deeper levels of the tree. The regret guarantees provided in the coming sections are naturally sensitive to this aspect. Their precise dependence on the tree structure can serve as a guiding metric when constructing the tree starting from a more general similarity model. In this work, we will mostly focus on the regret minimization aspect, taking the tree structure for granted. The application of Section 7 is the exception to this, as there, we also undertake the task of designing the tree.

Online bandit game and multi-point feedback We now describe the sequential protocol and the regret notion used throughout the paper. Fix an integer m between 1 and $L+1$. At every round t, the environment secretly draws a tree-compatible loss function y_t, and simultaneously, the learner plays $m \geq 1$ actions $A_{t,1}, \dots, A_{t,m} \in \mathcal{A}$, observes $y_t(A_{t,1}), \dots, y_t(A_{t,m})$, and incurs their average $\frac{1}{m}\sum_{i=1}^{m} y_t(A_{t,i})$. The regret at horizon T is

$$\mathbf{Reg}_{T,m} := \mathbb{E}\left[\sum_{t=1}^{T} \frac{1}{m}\sum_{i=1}^{m} y_t\big(A_{t,i}\big)\right] - \min_{a\in\mathcal{A}} \mathbb{E}\left[\sum_{t=1}^{T} y_t(a)\right],$$

which the learner aims to minimize. Note that the case $m = 1$ coincides with the standard (one-point) bandit setting. The case of $m \geq 2$ is considered in the light of the impossibility result presented in Section 3. We assume that y_t is independent of the learner's actions at round t or later, conditioned on the random events up to the end of round $t-1$. Besides this, we make no further assumption about the loss generating mechanism. However, in Sections 5 and 6 we achieve improved guarantees when the loss sequence is benign in the following sense.

Condition 1 (Generalized Stochastic Setting) *There exists a vector $\Delta \in \mathbb{R}_+^K$ satisfying $|\{a \in \mathcal{A}\colon \Delta(a) = 0\}| \leq 1$ such that $\mathbf{Reg}_{T,m} \geq \mathbb{E}\sum_{t=1}^{T}\frac{1}{m}\sum_{i=1}^{m}\Delta(A_{t,i})$ holds for any algorithm.*

Conditions of this form are common in the literature (see e.g., Zimmert and Seldin (2021); Ito et al. (2022); Dann et al. (2023)), in particular, this subsumes the standard stochastic setting where the losses are drawn from a fixed distribution at each round.

3. Lower Bound: The Insufficiency of One-Point Feedback

In the absence of any structure, the minimax regret of the bandit problem is $\Theta(\sqrt{KT})$ (Auer et al., 2002). We now show that also for our problem, an improvement of this rate is *information-theoretically impossible* under one-point bandit feedback: no matter how small the σ values are, the worst-case regret remains $\Omega(\sqrt{KT})$.

Fix a depth-L tree with leaf set $\mathcal{A}$ and scale parameters $(\sigma_j)_{j\in[L]}$. Define the class

$$\mathcal{Y}_\sigma^T := \Big\{(y_t)_{t=1}^T : \forall t,\ y_t : \mathcal{A} \to [0,1],\ \forall a,b \in \mathcal{A},\ |y_t(a) - y_t(b)| \leq \sigma_{s(a,b)}\Big\}.$$

Recalling that $\sigma_L = \min_{j\in[L]}\sigma_j$, we now formally state the lower bound.

Proposition 1 (One-point lower bound under tree-induced similarity) *Let $\rho \coloneqq \sigma_L$. For $K > 2$, $T > 32(K-1)\log(14)$, and $\rho > 0.22\sqrt{(K-1)/T}$, any randomized one-point bandit algorithm satisfies*

$$\sup_{(y_t)_{t=1}^T \in \mathcal{Y}_\sigma^T} \mathbb{E}\big[\mathbf{Reg}_{T,1}(y_{1:T})\big] \;>\; \frac{1}{504}\sqrt{T(K-1)}.$$

Equivalently, for any such algorithm there exists a tree-compatible loss sequence $(y_t)_{t=1}^T \in \mathcal{Y}_\sigma^T$ inducing $\Omega(\sqrt{KT})$ expected regret.

Proof sketch. The argument is a direct reduction to the range-impossibility result of Gerchinovitz and Lattimore (2016, Corollary 4). Let

$$C_\rho \coloneqq \Big\{x \in [0,1]^K : \max_{a,b\in\mathcal{A}} |x(a) - x(b)| \le \rho\Big\} \qquad \text{and} \qquad C_\rho^T \coloneqq \Big\{(y_t)_{t=1}^T : \forall t,\ y_t \in C_\rho\Big\}.$$

Any $x \in C_\rho$ satisfies $|x(a) - x(b)| \le \rho = \sigma_L \le \sigma_{s(a,b)}$, hence $C_\rho^T \subseteq \mathcal{Y}_\sigma^T$. Applying (Gerchinovitz and Lattimore, 2016, Corollary 4) with $\rho = \sigma_L$ yields the claimed $\Omega(\sqrt{KT})$ bound over $\mathcal{Y}_\sigma^T$. ■

Intuition and impact on the sequel. A convenient way to interpret C_ρ is via a "baseline + perturbation" representation: each $x \in C_\rho$ can be written as $x(a) = b + u(a)$ where $b = \min_a x(a) \in [0, 1-\rho]$ and $u(a) = x(a) - b \in [0, \rho]$. With one-point feedback, the learner observes only a single absolute value $y_t(A_t) = b_t + u_t(A_t)$ per round and cannot disentangle the unknown baseline b_t from the informative part $u_t(A_t)$; an adversary may drift b_t over time without violating the range constraint, making the observations informationally ambiguous even when ρ is tiny. This motivates our focus on multi-point feedback: contrasts such as $y_t(A_{t,1}) - y_t(A_{t,2})$ cancel the common baseline and expose the effective range, making the structure of the action set exploitable.

4. Basic Results for the Standard Bandit Problem

To properly contextualize the techniques used in the coming sections, we lay out here a brief summary of some basic tools and results pertaining to the standard bandit problem. A standard approach for tackling said problem is through a reduction to online convex optimization, employing an algorithm like follow-the-regularized-leader (FTRL) or mirror descent over the probability simplex of the actions as a decision set, see, e.g., (Orabona, 2025). In particular, if we fix a sequence $(z_t)_{t\ge1}$ of vectors in $\mathbb{R}^K$ and a sequence of regularizers $(\psi_t)_{t\ge1}$ with $\psi_t \colon \mathbb{R}^K \to \mathbb{R}$ being a convex function whose domain includes Δ_K, the FTRL algorithm produces a sequence of distributions over actions $(p_t)_{t\ge1}$ given by

$$p_1 = \operatorname*{arg\,min}_{p\in\Delta_K} \psi_1(p) \quad \text{and} \quad p_t = \operatorname*{arg\,min}_{p\in\Delta_K} \big\langle \textstyle\sum_{s\le t-1} z_s, p\big\rangle + \psi_t(p) \;\text{ for } t \ge 2\,,$$

which one can utilize by sampling A_t from p_t. A standard choice for $z_t(a)$ is $[\![a = A_t]\!] y_t(a)/p_t(a)$, which is an unbiased estimate of $y_t(a)$. The last ingredient to specify is the regularizer ψ_t. A common form in bandit problems is $\psi_t(p) \coloneqq (1/\eta_t)\sum_{a\in\mathcal{A}} f(p(a))$ for $p \in \Delta_K$, where f is a one-dimensional convex function. Assuming f is Legendre (see Rockafellar 1970, Chapter 26), it holds for any $q \in \Delta_K$ that

$$\begin{aligned}\textstyle\sum_{t=1}^T \langle z_t, p_t - q\rangle \le \psi_{T+1}(q) - \psi_1(p_1) + \textstyle\sum_{t=1}^T \psi_t(p_{t+1}) - \psi_{t+1}(p_{t+1}) \\ + \textstyle\sum_{t=1}^T (1/\eta_t)\sum_{a\in\mathcal{A}} D_{f^*}\big(f'(p(a)) - \eta_t(z_t(a) - y_t(A_t))\big\| f'(p(a))\big)\,, \quad (2)\end{aligned}$$

where D_{f^*} is the Bregman divergence induced by the convex conjugate of f. Note that the expectation of the left-hand side is the regret of the learner that samples A_t from p_t. The terms in the first line of the bound are collectively referred to as the bias term, which depends primarily on the magnitude of the regularizer; while the last sum is usually referred to as the variance term, this is a measure of the stability of the sequence $(p_t)_{t\geq 1}$, which depends on the magnitude of the losses and the curvature of the regularizer.

Setting $\eta_t = \eta/\sqrt{t}$ with $\eta > 0$ and $f(x) = x \ln x$ recovers the well-known EXP3 algorithm, whose variance and bias terms are respectively bounded in expectation by $\eta K \sum_{t=1}^{T} 1/\sqrt{t}$ and $(1/\eta)\mathbb{E}\sum_{t=1}^{T}(1/\sqrt{t})\sum_a p_t(a)\ln(1/p_t(a))$ up to constants, assuming $y_t \in [0,1]^K$. A suitable tuning of η yields a regret bound of order $\sqrt{KT\ln K}$. Note that the regularizer used by EXP3 is the negative Shannon entropy function. Alternatively, one can choose $f(x) = 2(x - \sqrt{x})$, which yields the negative $1/2$-Tsallis entropy as the regularizer. Using this function with $\eta_t = 1/\sqrt{t}$, the bias and variance terms are nicely balanced, both being bounded in expectation by $\mathbb{E}\sum_{t=1}^{T}(1/\sqrt{t})\sum_a(\sqrt{p_t(a)} - p_t(a))$ up to constants. This yields an improved regret bound of order $\sqrt{KT}$, which is unimprovable in the minimax sense (Auer et al., 2002).

Besides a minor improvement in the worst case regime, this algorithm enjoys a major advantage against easy loss sequences. In particular, since its regret scales with the sum of the Tsallis entropy of p_t at each round, one expects improved performance if the sequence $(p_t)_{t\geq 1}$ rapidly concentrates mass on a single action. One such scenario is when the losses are drawn at every round in an i.i.d. manner from a fixed distribution. More generally, under Condition 1, the algorithm enjoys a regret bound of order $\sum_{a:\,\Delta(a)>0}\ln(T)/\Delta(a)$, see (Zimmert and Seldin, 2021). It is worth noting that one can also achieve a logarithmic regret bound in the stochastic setting by equipping EXP3 with a certain adaptive sequence of learning rates $(\eta_t)_{t\geq 1}$, even if it is an inferior bound of order $K\ln(T)^2/\min_{a:\,\Delta(a)>0}\Delta(a)$, see (Ito et al., 2022).

5. Warm-up: Full-Path Feedback

In this section, we adopt the more structured model of Martin et al. (2022). The purpose here is to illustrate our results, specifically, our best-of-both-worlds bounds, in a simpler setting and provide an algorithmic template to build upon in the coming sections. To recall, compared to ours, the setting of Martin et al. (2022) involves two interdependent restrictions. The first, posits that the adversary selects at every round a function $\ell_t\colon \mathcal{V} \to [0,1]$ assigning a loss to every node in the tree such that $\ell_t(v) \leq R_j$ for all $v \in V_j$, where $(R_j)_{j\in[L]}$ are fixed positive weights in $[0,1]$. The loss $y_t(a)$ of any action $a \in \mathcal{A}$ is then taken as $\sum_{v\succ a}\ell_t(v)$. Note that our level-wise smoothness property can be imposed on this model taking $R_j = \sigma_j - \sigma_{j+1}$, though the class of admissible loss functions is now smaller. The second and more crucial restriction is that the learner gets to observe all the values $(\ell_t(v))_{v\succ a}$ upon choosing $A_t = a$. In practical scenarios, it can be difficult to specify the ℓ_t function and unrealistic to assume we can access its values along the whole path from the root to the played action. Nevertheless, we entertain this model for the moment as a starting point, providing in the process improved results compared to Martin et al. (2022).

In the worst case (adversarial) regime, Martin et al. (2022) showed that it is possible to leverage the similarity structure of the problem and the extra feedback to improve upon the minimax optimal $\sqrt{KT}$ regret rate of the standard bandit problem. In particular, they achieve a bound of order $\sqrt{K_{\mathrm{eff}}(R)T\log K}$, where $K_{\mathrm{eff}}(R) \coloneqq \left(\sum_{j\in[L]} R_j\sqrt{|V_j|}\right)^2$ is seen as the effective number of actions. When $\sum_{j\in[L]} R_j = 1$, as assumed by Martin et al. (2022), K_{eff} is never worse than

K, and can be much smaller when the tree has a favorable structure, an example is provided at the end of this section. The algorithm they use to achieve this result is a nested version of the EXP3 algorithm, which can be seen as an instance of FTRL with two main features that distinguish it from the examples provided in the previous section. Firstly, they leverage the semi-bandit feedback by using $\widehat{\ell}_t(v) \coloneqq [\![v \succ A_t]\!]\ell_t(v)/p_t[v]$ as an unbiased estimate of $\ell_t(v)$, leading to the choice of $z_t(a) = \sum_{v \succ a} \widehat{\ell}_t(v)$.[1] The second feature is the use of a nested Shannon entropy regularizer defined as $\psi_t(p) = (1/\eta_t) \sum_{j \in [L]} \alpha_j \sum_{v \in V_j} p[v] \ln p[v]$, which is a weighted sum (with non-negative weights $(\alpha_j)_{j \in [L]}$) of the Shannon entropy of the aggregated version of p at every level.

Ultimately though, the use of this special form of regularizer is not fully justified by the results of Martin et al. (2022). In particular, they present an upper bound of order $(K_{\mathrm{eff}}(R)/\alpha_L) \sum_{t=1}^T \eta_t$ on the variance term of their algorithm, which holds even if $\alpha_1 = \cdots = \alpha_{L-1} = 0$, i.e., the standard version of EXP3. Their bound on the bias term takes a complex form that does involve all the weights, but is less significant as its dependence on the number of actions is logarithmic. To gain more insight, we show that for any regularizer of the form $\psi_t(p) = (1/\eta_t) \sum_{j \in [L]} \alpha_j \sum_{v \in V_j} f(p[v])$ with f being Legendre, the bound in (2) holds with the last sum replaced by (see Lemma 11)

$$\sum_{t=1}^T 1/\eta_t \sum_{j \in [L]} \alpha_j \sum_{v \in V_j} D_{f^*}\big(f'(p[v]) - \eta_t \zeta_t(v)/\alpha_j \big\| f'(p[v])\big)\,,$$

where $\zeta_t \colon \mathcal{V} \to \mathbb{R}$ is any function satisfying $\sum_{v \succ a} \zeta_t(v) - \sum_{w \succ a'} \zeta_t(w) = z_t(a) - z_t(a')$ for all $a, a' \in \mathcal{A}$. When $f(x) = x \ln x$, one can show that choosing $\zeta_t(v) = [\![v \in \mathcal{A}]\!] z_t(a)$, i.e., pushing all the weights to the last level, allows recovering the aforementioned bound of $(K_{\mathrm{eff}}(R)/\alpha_L) \sum_{t=1}^T \eta_t$. It is unclear if this can be significantly improved upon; moreover, one can safely neglect the tradeoff with the bias term (considering the tuning of the α weights) due to the great disparity of magnitude between bias and variance in this case (logarithmic vs linear in the number of actions).

On the other hand, picking $f(x) = 2(x - \sqrt{x})$ and making the more natural choice of $\zeta_t(v) = \widehat{\ell}_t(v)$ (up to a constant shift) balances the bias and variance bounds, as long as $\alpha_j \propto R_j$. Specifically, one obtains a bound of order $\sum_{t=1}^T 1/\sqrt{t}\, \mathbb{E} \sum_{j \in [L]} R_j \sum_{v \in V_j} \sqrt{p_t[v]} - p_t[v]$, which is an analogue of the entropy-dependent bound discussed in the previous section for the unstructured bandit problem. The advantage of this result is, again analogously to the discussion in the previous section, that it can be used to obtain improved bounds in more benign scenarios where the losses are drawn at every round from a fixed distribution, which was not addressed by Martin et al. (2022). Crucially, the appearance of the scale parameters in this bound allows leveraging the structure of the problem to obtain improved bounds (compared to unstructured bandits) also in the stochastic case. Imperative to this is the nested form of the regularizer, which injects the scale parameters also into the bias term. We present then in the following theorem a best-of-both-worlds: a slight improvement of the worst-case bound of Martin et al. (2022) and a new structure-aware logarithmic regret bound holding in the stochastic setting. Recall that $K_{\mathrm{eff}}(R) \coloneqq \big(\sum_{j \in [L]} R_j \sqrt{|V_j|}\big)^2$, and define

$$K_{\mathrm{eff},\Delta}(R) \coloneqq \big(\textstyle\sum_{j \in [L]} R_j \sqrt{\Gamma(j,\Delta)}\big)^2 \quad \text{and} \quad \Gamma(j,\Delta) \coloneqq \displaystyle\sum_{v \in V_j \colon \min_{a \prec v} \Delta(a) \neq 0} \frac{1}{\min_{a \prec v} \Delta(a)}\,.$$

Theorem 2 *FTRL with* $\psi_t(p) = \sqrt{5 \max\{t, 20\}} \sum_{j \in [L]} R_j \sum_{v \in V_j} p[v] - \sqrt{p[v]}$ *and* $z_t(a) = \sum_{v \succ a} [\![v \succ A_t]\!] \ell_t(v)/p_t[v]$ *satisfies*

$$\mathbf{Reg}_{T,1} \leq \begin{cases} 15\sqrt{K_{\mathrm{eff}}(R)} + 4\sqrt{5 K_{\mathrm{eff}}(R) T} \\ 30\sqrt{K_{\mathrm{eff}}(R)} + 20 K_{\mathrm{eff},\Delta}(R) \ln(eT) \quad \textit{(under Condition 1)} \end{cases}$$

1. We use $[\![\cdot]\!]$ to denote the indicator function.

The proof can be found in Appendix C. For comparison, algorithms like UCB and Tsallis-INF enjoy in the stochastic setting a regret bound of order $\sum_{a:\,\Delta(a)>0} \ln(T)/\Delta(a)$, which is not improvable in general in the unstructured bandit problem. Defining $K_\Delta \coloneqq \sum_{a:\,\Delta(a)>0} 1/\Delta(a)$, it is easy to see that $K_{\mathrm{eff},\Delta}(R)$ improves upon K_Δ, assuming as before that $\sum_{j=1}^{L} R_j = 1$. In particular, Jensen's inequality gives that $K_{\mathrm{eff},\Delta}(R) \leq \sum_{j\in[L]} R_j \Gamma(j,\Delta)$, and clearly, $\Gamma(j,\Delta) \leq K_\Delta$ at any level j since $|V_j| \leq K$ and every action has a unique ancestor in V_j. In fact, it is usually strictly better; we demonstrate now that $K_{\mathrm{eff},\Delta}(R)$ may be exponentially smaller than K_Δ on a simple example with $\mathcal{A} \subset [0,1]$ and a dyadic partitioning of the space that correlates the losses of the actions. Let $L \geq 1$ and consider for each level $V_j := \{i2^{-j}; i = 0,\dots,2^j-1\}$, $\mathcal{A} := V_L$ and $K = 2^L$. For each node $v \in V_j$, we set $\mathrm{prt}(v) = \lfloor v2^{j-1} \rfloor 2^{1-j}$and $\ell_t(v) = 0.5(v - \mathrm{prt}(v)) + 2^{-(j+1)}\varepsilon_t(v)$ with $\varepsilon_t(v) \sim$ Bernoulli$(1/2)$ (we assign a loss 0 to the root). This implies that $\ell_t(v) \in [0, 2^{-j}]$ and $\min_{a\prec v} \Delta(a) = v/2$. Then, for all $j \in [L]$, $\Gamma(j,\Delta) = \sum_{i=1}^{2^j-1} \frac{2^{j+1}}{i} \leq (j+1)2^{j+1}$, which, using $R_j \leq 2^{-j}$, yields $K_{\mathrm{eff}}(R) = \left(\sum_{j\in[L]} 2^{-\frac{j}{2}}\right)^2 \leq 6$ and

$$K_{\mathrm{eff},\Delta}(R) \coloneqq \left(\sum_{j\in[L]} R_j \sqrt{\Gamma(j,\Delta)}\right)^2 \leq \left(\sum_{j\in[L]} 2^{\frac{1-j}{2}} \sqrt{j+1}\right)^2 \leq 6(L+1) = 12(\log_2 K + 1).$$

On the other hand, we have $K_\Delta = \Gamma(L,\Delta) \geq 2^{L+1} L \ln 2 = 2K \ln K$.

6. Multi-Point Bandit Feedback

Having seen that exploiting the structure of our problem is impossible under (one-point) bandit feedback, we seek to demonstrate in this section that one can skirt this obstacle by executing more than one action in each round and observing all their losses. Precisely, just making two observations per round is already sufficient, but in some case observing more (up to $L+1$) can lead to better performance. Importantly, the learner does not make these observations for free, but pays the average loss of the executed actions (some of which can be duplicates). This is decidedly milder than the feedback model considered in Section 5; not least of all because the environment (its feedback mechanism in particular) is not strongly coupled to the structure of the tree.

Nevertheless, the manner in which we exploit this limited feedback is to simulate the full-path feedback of Section 5 and use the same algorithmic template. There, we leveraged the fact that at every round, there exists a loss function $\ell_t \colon \mathcal{V} \to \mathbb{R}$ such that **(i)** $\sum_{v\succ a} \ell_t(v) = y_t(a)$ for all $a \in \mathcal{A}$; **(ii)** one observes $(\ell_t(v))_{v\succ A_t}$ after executing A_t; and **(iii)** $\ell_t(v) \leq R_j$ for all $v \in V_j$. Here, however, we will build our own ℓ_t function satisfying similar properties. Firstly, we associate with every node $v \in \overline{\mathcal{V}}$ an action $\omega_t(v)$ that is sampled from p_t conditioned on having v as an ancestor. Then, define

$$\ell_t(v) \coloneqq y_t(\omega_t(v)) - y_t(\omega_t(\mathrm{prt}(v)))\,.$$

Now, it is easy to see that Property (**i**) holds up to a constant shift (by $y_t(w_t(v_0))$), which suffices. Moreover, an analogue of Property (**iii**) holds via the loss smoothness condition (1); in particular, $|\ell_t(v)| \leq \sigma_j$ for all $v \in V_j$. Finally, for Property (**ii**), we sample an action $A_{t,1} \sim p_t$ and execute it along with $\omega_t(v)$ for all ancestors v of $A_{t,1}$ (including v_0), thus we observe $(\ell_t(v))_{v\succ A_{t,1}}$. Note that in this manner, we execute $L+1$ actions, the marginal law of each is p_t, conditioned on the events of past rounds. In practice, of course, one only needs to sample $\omega_t(\cdot)$ at the ancestors of $A_{t,1}$.

Using this reduction, we can recover the bounds of Theorem 2, but now for $\mathbf{Reg}_{T,L+1}$ and with R swapped for σ. Going a step further, at a slight cost, we can get away with observing ℓ_t at only one node, which we can achieve by executing just two actions. In particular, we sample a level $j_t \in [L]$

Input: sequence of regularizers $(\psi_t)_{t\geq 1}$, sequence of constants $(b_j)_{j=1}^L$, distribution $\delta \in \Delta_L$
Initialization: $p_1 = \arg\min_{p\in\Delta_K} \psi_1(p)$
for $t \geq 1$ **do**
  sample $A_{t,1} \sim p_t$, $j_t \sim \delta$, and $A_{t,2} \sim p_t \mid A_{t,2} \sim_{j_t-1} A_{t,1}$
  for all $a \in \mathcal{A}$, set $z_t(a) = [\![a \sim_{j_t} A_{t,1}]\!] \dfrac{y_t(A_{t,1}) - y_t(A_{t,2}) + b_{j_t}}{\delta_{j_t} p_t[\mathrm{anc}_{j_t}(A_{t,1})]}$
  set $p_{t+1} = \arg\min\limits_{p\in\Delta_K} \langle \sum_{s\leq t} z_s, p\rangle + \psi_{t+1}(p)$
end

Algorithm 1: FTRL with 2-Point Bandit Feedback

from a fixed distribution δ and a node v_t with probability $\propto [\![\cdot \in V_{j_t}]\!] p_t[\cdot]$. Then, we simply execute $\omega_t(v_t)$ and $\omega_t(\mathrm{prt}(v_t))$, giving us access to $\ell_t(v_t)$. A more direct version of this procedure is given in Algorithm 1. Note that we account for the reduced feedback by correcting the loss estimate with a factor of $1/\delta_{j_t}$. Additionally, we shift ℓ_t by a level-dependent constant to enforce non-negativity for technical reasons. Ultimately, we arrive at the following theorem (see Appendix D).

Theorem 3 *Algorithm 1 with* $\psi_t(p) = 2\sqrt{6}\sum_{j\in[L]}(\sigma_j/\sqrt{\delta_j})\sqrt{\max\{t, 6/\delta_j\}}\sum_{v\in V_j} x[v] - \sqrt{x[v]}$ *and* $b_j = \sigma_j$ *for all* $j \in [L]$ *satisfies*

$$\mathbf{Reg}_{T,2} \leq \begin{cases} 18\sqrt{K_{\mathrm{eff}}(\sigma/\delta)} + 8\sqrt{6}\sqrt{K_{\mathrm{eff}}(\sigma/\sqrt{\delta})T} \\ 36\sqrt{K_{\mathrm{eff}}(\sigma/\delta)} + 96K_{\mathrm{eff},\Delta}(\sigma/\sqrt{\delta})\ln(eT) \quad \textit{(under Condition 1)} \end{cases}.$$

We used here the notation of Theorem 2, specifically, $K_{\mathrm{eff}}(\sigma/\sqrt{\delta}) \coloneqq \left(\sum_j (\sigma_j/\sqrt{\delta_j})\sqrt{|V_j|}\right)^2$ and $K_{\mathrm{eff},\Delta}(\sigma/\sqrt{\delta}) \coloneqq \left(\sum_j (\sigma_j/\sqrt{\delta_j})\sqrt{\Gamma(j,\Delta)}\right)^2$. The choice of δ was left open in the bounds above. At least for the adversarial regime, one can choose δ so as to minimize $K_{\mathrm{eff}}(\sigma/\sqrt{\delta})$. In favorable cases, carefully choosing δ can render its effect on the bound negligible, see Section 7. In general, choosing the uniform distribution is a safe choice that results in the bound scaling with an extra factor of $\sqrt{L}$ in the adversarial regime and L in the stochastic one. If one has access to m-point feedback, where $2 \leq m \leq L+1$, then Algorithm 1 can be generalized by randomly drawing $m-1$ levels at every round. We show in Appendix D that the bounds of Theorem 3 continue to hold in that case, with δ_j denoting the probability that level j is amongst the chosen levels at any round. Drawing the $m-1$ levels uniformly at random without replacement results in the bounds scaling with $\sqrt{L/(m-1)}$ and $L/(m-1)$ in the adversarial and the stochastic regimes respectively.

The issue of δ aside, the utility of the bounds presented here (compared to standard bandit bounds) relies on the tree being carefully constructed; specifically, that the values of σ_j decay quickly relative to the increase in $|V_j|$. For instance, in the example presented in Section 5, $\sigma_j \leq 2^{-j+1}$, hence the bounds derived there continue to hold up to constant factors, illustrating again the huge advantage that K_{eff} and $K_{\mathrm{eff},\Delta}$ can bring. In general, these parameters are to be used as an efficiency criteria when constructing a tree from a similarity structure. We provide more insights on this issue in the next section.

7. Application: Lipschitz Bandits with Two-Point Feedback

Under the same interaction model as before, suppose we are now given an arbitrary action set $\mathcal{A}$ and a set $\mathcal{Y}$ of admissible loss functions, whose members map $\mathcal{A}$ to $\mathbb{R}$. We can always equip $\mathcal{A}$ with

the following metric:

$$d_{\mathcal{A},\infty}(a,b) \coloneqq \sup_{y\in\mathcal{Y}} |y(a) - y(b)| \;\; \forall a,b \in \mathcal{A}\,. \tag{3}$$

To be precise, this function is a metric only if no two actions behave identically on all loss functions in $\mathcal{Y}$, which is fair to impose. All members of $\mathcal{Y}$ are obviously 1-Lipschitz with respect to $d_{\mathcal{A},\infty}$.

In the tree-based model we have studied thus far, $\mathcal{Y}$ was taken as the set of tree-compatible loss functions described in Section 2. In particular, the induced metric in this case is given by $d_{\mathcal{A},\infty}(a,b) = \sigma_{s(a,b)}$, where, to recall, $s(a,b) \coloneqq 1 + \max\{j \in \{0,\dots,L\} \mid a \sim_j b\}$. More than a metric, this function satisfies a strong form of the triangle inequality: $\sigma_{s(a,b)} \leq \max\{\sigma_{s(a,c)}, \sigma_{s(c,b)}\}$ for all $a,b,c \in \mathcal{A}$ since $\sigma_1 \geq \cdots \geq \sigma_L$; hence, it is an *ultrametric*. Note that the tree structure can be recovered from this distance function if the similarity scales are strictly decreasing. A similar notion of distance was studied by Slivkins (2011) in the stochastic setting, and was referred to as *implicit distance*. Seen from this angle, the algorithmic techniques presented so far are designed to leverage this special form of distance and the implied restriction on the losses via (3).

We aim to illustrate here that our results can still be adapted to cases in which $d_{\mathcal{A},\infty}$ is a generic metric (or pseudometric). This is done, naturally, by constructing a tree over the action set and forcing a specific decay schedule for the resulting similarity scales. This constitutes a "truncation" or "compression" of the original metric structure, but still leads to interesting results. We will focus henceforth on the two-point feedback setting. In fact, the arguments used to prove the one-point feedback lower bound of Proposition 1 can be easily adapted to show that it still holds in this more general setting provided that $\mathcal{A}$ is finite and the conditions of the proposition hold with $\rho = \min_{a,b\in\mathcal{A}\colon a\neq b} d_{\mathcal{A},\infty}(a,b)$.

Towards employing Algorithm 1 and its regret bound, we start by constructing a tree whose nodes are members of $\mathcal{A}$. Unlike before, the nodes in the last level of this tree need not include the entire set $\mathcal{A}$ (which we do not require to be finite here). We specify now two parameters to be used in the sequel: a constant $c > 0$ and a positive integer L representing the number of levels as before. We will assume that $c \geq \max_{a,b\in\mathcal{A}} d_{\mathcal{A},\infty}(a,b)$. For $j \in [L]$, let $\mathcal{A}_j \subseteq \mathcal{A}$ be a subset of actions that form a $c2^{-j}$ covering with respect to $d_{\mathcal{A},\infty}$.[2] For convenience, let $\mathcal{A}_0 \coloneqq \{a_0\}$ with some arbitrary action a_0. Then, for any $a \in \mathcal{A}_j$, define its parent as its closest action in $\mathcal{A}_{j-1}$ with ties broken arbitrarily; that is, $\mathrm{prt}(a) \in \arg\min_{a'\in\mathcal{A}_{j-1}} d_{\mathcal{A},\infty}(a,a')$ with ties broken arbitrarily. If one action belongs to multiple covers, each of these occurrences is to be regarded as a distinct node.

In this manner, it is possible that certain nodes in intermediate levels end up with no children. Hence, we define a pruned version of the covering at each level j:

$$\overline{\mathcal{A}}_j = \{a \in \mathcal{A}_j \mid \exists a' \in \mathcal{A}_L \colon a = \mathrm{anc}_j(a')\}\,,$$

implying that $\overline{\mathcal{A}}_j \subseteq \mathcal{A}_j$ for each j and that $\overline{\mathcal{A}}_L = \mathcal{A}_L$. Now the sequence $(\overline{\mathcal{A}}_j)_{j=0,\dots,L}$ forms a tree, which we will refer to as $\mathcal{T}\big((\overline{\mathcal{A}}_j)_{j=0}^{L}\big)$, whose leaves are all at level L. This tree induces a nested sequence of partitions of the set $\mathcal{A}_L$, which will be the action set in the eyes of the algorithm. For

2. This means that $\min_{a'\in\mathcal{A}_j} d_{\mathcal{A},\infty}(a,a') \leq c2^{-j}$ for all $a \in \mathcal{A}$.

any pair of actions $a, b \in \mathcal{A}_L$ such that $s(a,b) = j$,[3] it holds that

$$\begin{aligned} d_{\mathcal{A},\infty}(a,b) &\le d_{\mathcal{A},\infty}(a, \mathrm{anc}_{j-1}(a)) + d_{\mathcal{A},\infty}(\mathrm{anc}_{j-1}(b), b) \\ &\le \sum_{k=j-1}^{L-1} d_{\mathcal{A},\infty}(\mathrm{anc}_{k+1}(a), \mathrm{anc}_k(a)) + \sum_{k=j-1}^{L-1} d_{\mathcal{A},\infty}(\mathrm{anc}_{k+1}(b), \mathrm{anc}_k(b)) \\ &\le 2c \sum_{k=j-1}^{L-1} 2^{-k} \le c 2^{3-j} \end{aligned}$$

where we have used the triangle inequality and the fact that $a \sim_{j-1} b$. Hence, via (3), setting $\sigma_j = c2^{3-j}$ is consistent with the assumption in (1), applied now to the new action set $\mathcal{A}_L$ and the tree $\mathcal{T}\big((\overline{\mathcal{A}}_j)_{j=0}^L\big)$. We can now instantiate the regret guarantee of Algorithm 1 on this tree structure, yielding the following result (the proof can be found in Appendix E). We will focus here on the adversarial setting.

Corollary 4 *Suppose we run Algorithm 1 on the tree $\mathcal{T}\big((\overline{\mathcal{A}}_j)_{j=0}^L\big)$ with parameters ψ_t, b_j as defined in Theorem 3, $\sigma_j = c2^{3-j}$, and $\delta_j \propto 2^{-2j/3}|\mathcal{A}_j|^{1/3}$. Then, assuming that $\min_{j\in[L]} \delta_j \ge 1/T$, it holds that*

$$\mathbf{Reg}_{T,2} \le 16 \cdot (9 + 4\sqrt{6}) \sqrt{K_{\mathrm{eff}}\big((\mathcal{A}_j)_{j=0}^L\big) T} + c\, 2^{-L} T\,,$$

where $K_{\mathrm{eff}}\big((\mathcal{A}_j)_{j=0}^L\big) \coloneqq c^2 \left(\sum_{j\in[L]} 2^{-2j/3}|\mathcal{A}_j|^{1/3}\right)^3$.

The last term is a bound on the approximation error resulting from using $\mathcal{A}_L$ as the action set. This holds because $\mathcal{A}_L$ is a $c2^{-L}$-cover of $\mathcal{A}$. This term can be turned into a small constant by choosing $L = \lfloor \log_2 T \rfloor$, though in some cases balancing this term with the first can be beneficial. When the cardinality of $\mathcal{A}$ is finite, denoted by K as before, then $K_{\mathrm{eff}}\big((\mathcal{A}_j)_{j=0}^L\big)$ is never larger than K up to small constants, regardless of the choice of L and the covers $(\mathcal{A}_j)_{j\in[L]}$. Of course, in interesting examples, ones where the covers can be made small at coarser levels, $K_{\mathrm{eff}}\big((\mathcal{A}_j)_{j=0}^L\big)$ can be made significantly smaller than K. Hence, one should strive to ensure that the covers are near-optimal, which can be achieved by a simple greedy allocation strategy. We elaborate more on this in Appendix E.2, where we also argue that no other tree-building strategy results, via the bound of Theorem 3, in a significantly smaller K_{eff}.

Now, let $\mathcal{A} \subset \mathbb{R}^d$ be a compact set with diameter D (w.r.t. to the Euclidean metric), and let $\mathcal{Y}$ be the set of G-Lipschitz loss functions from $\mathcal{A}$ to $\mathbb{R}$. It holds then that

$$d_{\mathcal{A},\infty}(a,b) \le G\|a - b\| \;\forall a, b \in \mathcal{A}\,.$$

In the ordinary interaction model—when the learner selects at each round a single action, observes the associated loss, and is evaluated through the regret $\mathbf{Reg}_{T,1}$— the minimax optimal regret rate is $\Omega(T^{(d+1)/(d+2)})$, as shown by Kleinberg et al. (2008). Subsequent works have focused on adapting to the effective (zooming) dimension d of the metric space (Slivkins, 2011; Kleinberg et al., 2019; Podimata and Slivkins, 2021), designing more adaptive algorithms (Héliou et al., 2021), or improving robustness to corruption (Kang et al., 2023). However, all these approaches rely on one-point

3. To recall, this means that the lowest common ancestor of a and b is at level $j-1$.

feedback and therefore suffer from the above minimax rate in the worst case. Under a richer feedback model ("one-sided" full-information) motivated by second-price auctions, Cesa-Bianchi et al. (2017) showed that this rate can be improved.

We show now that under two-point feedback, Corollary 4 can be used to obtain better regret guarantees compared to the one-point feedback minimax rate. We will optimize the depth of the tree depending on the dimension. For each $j = 1, \dots, L$, let $\varepsilon_j = D\sqrt{d}\,2^{-j}$ be a precision level, and let $\mathcal{A}_j$ be an ε_j-cover of $\mathcal{A}$ (w.r.t. to the Euclidean metric) with minimal cardinality. Using a uniform grid, one can show that $|\mathcal{A}_j| \leq (D\sqrt{d}/\varepsilon_j)^d = 2^{jd}$. Note that $\mathcal{A}_j$ is a $GD\sqrt{d}2^{-j}$-cover w.r.t. $d_{\mathcal{A},\infty}$; hence, we can choose $c = GD\sqrt{d}$, which is an upper bound for $\max_{a,b\in\mathcal{A}} d_{\mathcal{A},\infty}(a,b)$ as required. The following result is then easy to obtain from Corollary 4 (the proof can be found in Appendix E).

Corollary 5 *Let $T \geq 8$. Algorithm 1 run on the tree $\mathcal{T}\big((\overline{\mathcal{A}}_j)_{j=0}^L\big)$ induced by the covers described above with $L = \lfloor \frac{1}{d}\log_2(T) \rfloor$ and parameters ψ_t, b_j as defined in Theorem 3, $\sigma_j = GD\sqrt{d}\,2^{3-j}$ and $\delta_j \propto 2^{j(d-2)/3}$, satisfies the following regret upper bound*

$$\mathbf{Reg}_{T,2} \lesssim GD \cdot \begin{cases} \sqrt{T} & \text{if } d = 1 \\ (\ln T)^{3/2}\sqrt{T} & \text{if } d = 2 \\ \sqrt{d}T^{\frac{d-1}{d}} & \text{if } d > 2 \end{cases},$$

where $\lesssim$ denotes an inequality up to a universal constant.

Interestingly, when $d \leq 2$, the above result shows that two-point feedback is sufficient to recover (up to logarithmic factors for $d = 2$) the optimal $\Omega(\sqrt{T})$ full-information rate. This follows from the fact that the effective number of arms induced by the underlying tree structure is $O((\log T)^3)$ for $d \leq 2$, even though the total number of leaves can be as large as T.

Acknowledgments

This work was supported by funding from the French government, managed by the National Research Agency (ANR), under the France 2030 program, reference ANR-23-IACL-0006. Additionally, this work was supported in part by ANR in the framework of the PEPR IA FOUNDRY project (ANR-23-PEIA-0003). PM is also a member of the Archimedes Research Unit/Athena RC, and was partially supported by project MIS 5154714 of the National Recovery and Resilience Plan Greece 2.0 funded by the European Union under the NextGenerationEU Program.

References

Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. *Advances in neural information processing systems*, 24, 2011.

Alekh Agarwal, Ofer Dekel, and Lin Xiao. Optimal algorithms for online convex optimization with multi-point bandit feedback. In *Colt*, pages 28–40, 2010.

Shipra Agrawal and Navin Goyal. Thompson sampling for contextual bandits with linear payoffs. In *International conference on machine learning*, pages 127–135. PMLR, 2013.

Imad Aouali, Branislav Kveton, and Sumeet Katariya. Mixed-effect thompson sampling. In *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics*, volume 206 of *Proceedings of Machine Learning Research*, pages 2087–2115. PMLR, 2023.

Peter Auer, Nicolo Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47(2):235–256, 2002.

Dimitri Bertsekas. *Convex optimization theory*, volume 1. Athena Scientific, 2009.

Sébastien Bubeck, Nicoló Cesa-Bianchi, and Sham M. Kakade. Towards minimax policies for online linear optimization with bandit feedback. In *Proceedings of the 25th Annual Conference on Learning Theory*, volume 23 of *Proceedings of Machine Learning Research*, pages 41.1–41.14. PMLR, 2012.

Nicolò Cesa-Bianchi, Pierre Gaillard, Claudio Gentile, and Sébastien Gerchinovitz. Algorithmic chaining and the role of partial feedback in online nonparametric learning. In *Conference on Learning Theory*, pages 465–481. PMLR, 2017.

Richard Combes, Mohammad Sadegh Talebi Mazraeh Shahi, Alexandre Proutiere, et al. Combinatorial bandits revisited. *Advances in neural information processing systems*, 28, 2015.

Richard Combes, Stefan Magureanu, and Alexandre Proutiere. Minimal exploration in structured stochastic bandits. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

Chris Dann, Chen-Yu Wei, and Julian Zimmert. A blackbox approach to best of both worlds in bandits and beyond. In *Proceedings of Thirty Sixth Conference on Learning Theory*, volume 195 of *Proceedings of Machine Learning Research*, pages 5503–5570. PMLR, 12–15 Jul 2023.

Rémy Degenne, Evrard Garcelon, and Vianney Perchet. Bandits with side observations: Bounded vs. logarithmic regret. In *Conference on Uncertainty in Artificial Intelligence*, 2018.

John C Duchi, Michael I Jordan, Martin J Wainwright, and Andre Wibisono. Optimal rates for zero-order convex optimization: The power of two function evaluations. *IEEE Transactions on Information Theory*, 61(5):2788–2806, 2015.

Abraham D Flaxman, Adam Tauman Kalai, and H Brendan McMahan. Online convex optimization in the bandit setting: gradient descent without a gradient. *arXiv preprint cs/0408007*, 2004.

Sébastien Gerchinovitz and Tor Lattimore. Refined lower bounds for adversarial bandits. *Advances in Neural Information Processing Systems*, 29, 2016.

Amélie Héliou, Matthieu Martin, Panayotis Mertikopoulos, and Thibaud Rahier. Zeroth-order nonconvex learning via hierarchical dual averaging. In *International Conference on Machine Learning*, pages 4192–4202. PMLR, 2021.

Joey Hong, Branislav Kveton, Sumeet Katariya, Manzil Zaheer, and Mohammad Ghavamzadeh. Deep hierarchy in bandits. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 8833–8851. PMLR, 2022a.

Joey Hong, Branislav Kveton, Manzil Zaheer, and Mohammad Ghavamzadeh. Hierarchical bayesian bandits. In Gustau Camps-Valls, Francisco J. R. Ruiz, and Isabel Valera, editors, *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pages 7724–7741. PMLR, 2022b.

Shinji Ito, Shuichi Hirahara, Tasuku Soma, and Yuichi Yoshida. Tight first-and second-order regret bounds for adversarial linear bandits. *Advances in Neural Information Processing Systems*, 33: 2028–2038, 2020.

Shinji Ito, Taira Tsuchiya, and Junya Honda. Nearly optimal best-of-both-worlds algorithms for online learning with feedback graphs. In *Advances in Neural Information Processing Systems*, volume 35, pages 28631–28643. Curran Associates, Inc., 2022.

David S. Johnson. Approximation algorithms for combinatorial problems. *Journal of Computer and System Sciences*, 9(3):256–278, 1974. ISSN 0022-0000.

Yue Kang, Cho-Jui Hsieh, and Thomas Chun Man Lee. Robust lipschitz bandits to adversarial corruptions. *Advances in Neural Information Processing Systems*, 36:10897–10908, 2023.

Robert Kleinberg. Nearly tight bounds for the continuum-armed bandit problem. In *Advances in Neural Information Processing Systems*, volume 17. MIT Press, 2004.

Robert Kleinberg, Aleksandrs Slivkins, and Eli Upfal. Multi-armed bandits in metric spaces. In *Proceedings of the fortieth annual ACM symposium on Theory of computing*, pages 681–690, 2008.

Robert Kleinberg, Aleksandrs Slivkins, and Eli Upfal. Bandits and experts in metric spaces. *Journal of the ACM (JACM)*, 66(4):1–77, 2019.

Tor Lattimore and Csaba Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.

Zhou Lu, Qiuyi Zhang, Xinyi Chen, Fred Zhang, David Woodruff, and Elad Hazan. Adaptive regret for bandits made possible: Two queries suffice. *arXiv preprint arXiv:2401.09278*, 2024.

Matthieu Martin, Panayotis Mertikopoulos, Thibaud Rahier, and Houssam Zenati. Nested bandits. In *International Conference on Machine Learning*, pages 15093–15121. PMLR, 2022.

Nicolas Nguyen, Imad Aouali, András György, and Claire Vernade. Prior-dependent allocations for bayesian fixed-budget best-arm identification in structured bandits. In *Proceedings of The 28th International Conference on Artificial Intelligence and Statistics*, volume 258 of *Proceedings of Machine Learning Research*, pages 379–387. PMLR, 2025.

Francesco Orabona. A modern introduction to online learning. *arXiv preprint arXiv:1912.13213*, 2025.

Chara Podimata and Alex Slivkins. Adaptive discretization for adversarial lipschitz bandits. In *Conference on Learning Theory*, pages 3788–3805. PMLR, 2021.

R. Tyrrell Rockafellar. *Convex analysis*. Princeton Mathematical Series. Princeton University Press, 1970.

Amir Sani, Gergely Neu, and Alessandro Lazaric. Exploiting easy data in online optimization. *Advances in Neural Information Processing Systems*, 27, 2014.

Ohad Shamir. An optimal algorithm for bandit and zero-order convex optimization with two-point feedback. *Journal of Machine Learning Research*, 18(52):1–11, 2017.

Petr Slavík. A tight analysis of the greedy algorithm for set cover. *Journal of Algorithms*, 25(2): 237–254, 1997. ISSN 0196-6774.

Aleksandrs Slivkins. Multi-armed bandits on implicit metric spaces. *Advances in Neural Information Processing Systems*, 24, 2011.

Tobias Sommer Thune and Yevgeny Seldin. Adaptation to easy data in prediction with limited advice. *Advances in Neural Information Processing Systems*, 31, 2018.

Zheng Wen, Branislav Kveton, and Azin Ashkan. Efficient learning in large-scale combinatorial semi-bandits. In *International Conference on Machine Learning*, pages 1113–1122. PMLR, 2015.

Julian Zimmert and Yevgeny Seldin. Tsallis-inf: An optimal algorithm for stochastic and adversarial bandits. *Journal of Machine Learning Research*, 22(28):1–49, 2021.

Appendix A. Motivation: Similarity structures and related notions

In this appendix, we elaborate more on the motivating example given in the introduction. In particular, we give more details on how a tree structure can be induced from this general setting.

A.1. From traits to trees

Consider again a finite set of *alternatives* (or *actions*) indexed by $a \in \mathcal{A} = \{1, \dots, K\}$. Based on their innate traits, these alternatives can be clustered in different ways so that each subgroup shares a common attribute: in our running example from the introduction, a collection of cars could be categorized by brand into Fords, Chevrolets, Toyotas, etc.; by type into sedans, compacts, SUVs, and so forth; by powertrain into gas, hybrid, or electric; etc. To formalize this idea, a *trait* is defined to be a partition T of $\mathcal{A}$ into pairwise disjoint subsets—called *classes*—each corresponding to a different *instance* of said trait. Thus, in the catalogue of cars considered above, `type` and `brand` would represent different traits, while Ford and Toyota would be different instances of the trait `brand`.

Accordingly, to define a similarity structure on $\mathcal{A}$, we will assume that $\mathcal{A}$ is equipped with a family of traits $\mathscr{T} = \{T_i\}_{i=1}^n$. Of course, two or more traits can be combined, leading to clusters like "Ford compacts", "Toyota sedans", etc. Formally, the *combination* (or *logical conjunction*) of two traits $T, T' \in \mathscr{T}$ is defined to be the "composite trait" $T \wedge T' = \{C \cap C' : C \in T, C' \in T'\}$; in words, $T \wedge T'$ simply groups together all alternatives that combine an instance of T with an instance of T'. More generally, any subset $\mathcal{S}$ of $\mathscr{T}$ gives rise to a composite trait via the operation $\mathcal{S} \mapsto \texttt{combine}(\mathcal{S}) \equiv \bigwedge_{T \in \mathcal{S}} T$. Then, to quantify all this, we will associate to each subset $\mathcal{S}$ of $\mathscr{T}$ a non-negative *heterogeneity score* $\varsigma_{\mathcal{S}}$: the smaller this score, the more similar any two alternatives agreeing on the traits in $\mathcal{S}$ are. We naturally posit that $\varsigma_{\mathscr{T}} = 0$ and that $\varsigma_{\mathcal{S}} \leq \varsigma_{\mathcal{S}'}$ when $\mathcal{S}' \subseteq \mathcal{S}$.

Moving forward, when selecting an alternative, one may first focus on the type of a car before worrying about its brand, or one could do the reverse. To model this, we define a *priority ranking* on a family of traits $\mathscr{T}$ to be a strict total order $>$ (read: "*has higher priority than*"). Given a priority ranking, we can define a sequence $(\mathcal{S}_j)_{j=0}^n$, where $\mathcal{S}_0 = \varnothing$, $\mathcal{S}_n = \mathscr{T}$, and $\mathcal{S}_j$ is the set composed of the first j traits. We thus obtain a nested sequence of partitions $\texttt{combine}(\mathcal{S}_0), \dots, \texttt{combine}(\mathcal{S}_n)$, each successively refining its predecessor. This sequence is nothing but a *tree structure* on $\mathcal{A}$, and, taken together, the family of traits, heterogeneity scores and the priority rank comprise what we call a *similarity structure* on $\mathcal{A}$.

In reference to the tree compatible loss model described in Section 2, if $\mathcal{S}_j$ is the combination of traits associated with the j-th level of the tree, then σ_{j+1} is simply $\varsigma_{\mathcal{S}_j}$.[4] Hence, the similarity scales depend on the priority ranking, and can hence affect the regret bounds. In the next section, we provide more technical details involving the aforementioned process of constructing the tree, and show how the number of logically admissible trees can be reduced.

A.2. A logical structure for the traits

In this section, our aim is to elaborate on certain aspects of priority rankings and their compatibility with the innate structure of the family of traits.

4. The $+1$ increment is due to the way $s(\cdot, \cdot)$ is defined.

Binary relations between traits To that end, we begin by considering certain natural binary relations among a family of traits $\mathscr{T} = \{T_i\}_{i=1}^n$ on $\mathcal{A}$. The first such relation is that of *refinement*: given two traits $T, T' \in \mathscr{T}$, we say that T is *finer* than T'—and we write $T \preccurlyeq T'$—if every element of T is a subset of some element of T'. Another natural relation arises when alternatives are categorized according to different traits, in which case some traits might be subordinate to other more "primitive", higher-order traits that precede them in any chain of logical reasoning. For instance, the `model` of a car cannot be defined without implicitly specifying the car's `brand`: after all, an Impala is, by definition, a Chevrolet.

To capture this, we define a *logical structure* on a family of traits $\mathscr{T}$ as a binary relation $\vdash$ (read: "*requires*" or "*is a primitive of*") satisfying the following conditions for all $T, T', T'' \in \mathscr{T}$:

(a) *Irreflexivity:* $T \nvdash T$.

(b) *Transitivity:* If $T \vdash T'$ and $T' \vdash T''$, then $T \vdash T''$.

(c) *Consistency:* If $T \vdash T'$, then $T \preccurlyeq T'$.

In short, condition (a) states that a trait T is not a primitive of itself, while condition (b) requires that primitives of primitives of T are themselves primitives of T; thus, taken together, conditions (a) and (b) simply say that $\vdash$ is a strict partial order on $\mathscr{T}$. The last condition (consistency) posits that two alternatives that differ by a primitive of T also differ by T itself: for example, since `brand` is a primitive of `model`, a Toyota and a Chevrolet can't be the same `model`. [5] For posterity, we will write the set of primitives of $T \in \mathscr{T}$ as

$$\texttt{prim}(T) = \{T' \in \mathscr{T} : T \vdash T'\}, \tag{4}$$

For example, if $\mathscr{T} = \{\texttt{brand}, \texttt{color}, \texttt{model}, \texttt{type}\}$, the implied logical structure readily gives $\texttt{prim}(\texttt{model}) = \{\texttt{brand}, \texttt{type}\}$ (an Impala is *de facto* a Chevrolet sedan), and $\texttt{prim}(\texttt{type}) = \texttt{prim}(\texttt{brand}) = \texttt{prim}(\texttt{color}) = \varnothing$ (since none of these traits implicitly requires another). A family of traits $\mathscr{T}$ equipped with a logical structure $\vdash$ will be called here a *similarity structure* on $\mathcal{A}$, and we will denote it as $\Lambda \equiv \Lambda(\mathscr{T}, \vdash)$.

Combining traits. We now proceed to detail the procedure of combining two or more traits, leading to clusters like "Ford compacts", "Toyota sedans", etc. In this case, a priority ranking gives rise to a nested succession of "composite" traits, which can be seen as a *decision tree:* at each step an additional trait is added to the mix, resulting in finer and finer partitions of $\mathcal{A}$ until all traits are present. Formally, define the *combination* (or *logical conjunction*) of two traits $T, T' \in \mathscr{T}$ as

$$T \wedge T' = \{C \cap C' : C \in T, C' \in T'\}, \tag{5}$$

i.e., as the coarsest partition of $\mathcal{A}$ that refines both T and T'. Intuitively, this means that $T \wedge T'$ groups together all alternatives that combine an instance of T with an instance of T'.[6] Then, any

5. This implication does not go the other way round: in particular, if T is finer than T', T' need not be a primitive of T. For example, if $\mathcal{A} = \{$a blue Ford, a yellow Chevy, a red Chevy$\}$ and $T_{\texttt{color}}$ is a partition of $\mathcal{A}$ by `color` and $T_{\texttt{brand}}$ is a partition by `brand`, the partition $T_{\texttt{color}}$ is (strictly) finer than $T_{\texttt{brand}}$, even though `brand` is not a primitive of `color`: in general, a car's `color` does not determine its `brand`, even though it happens to do so for this *particular* set of cars. This shows that the prerequisite relation $\vdash$ is about the traits themselves as abstract entities, not about the set of alternatives embodying those traits.

6. For book-keeping reasons, traits that are obtained this way will be called *composite traits*, to differentiate them from their "primitive" counterparts $T \in \mathscr{T}$: for example, "Toyota sedan" is the instance of the composite trait "`type` and `brand`" which describes all cars that are of `brand` "Toyota" and `type` "sedan" (the two primitives).

subset $\mathcal{S}$ of $\mathscr{T}$ gives rise to a composite trait via the operation

$$\mathcal{S} \mapsto \texttt{combine}(\mathcal{S}) \equiv \textstyle\bigwedge_{T\in\mathcal{S}} T. \tag{6}$$

Under `combine`, the empty set $\varnothing$ is mapped to the combination of *no* traits, i.e., the partition of $\mathcal{A}$ that refines no other partition, namely $\texttt{combine}(\varnothing) = \bigwedge_{T\in\varnothing} T \equiv 1$. Then, including more traits in $\mathcal{S}$ makes the resulting partition $\texttt{combine}(\mathcal{S})$ finer and finer, until there are no more traits left to tally. Hence, the finest partition induced by a family of traits $\mathscr{T} = \{T_i\}_{i=1}^n$ is

$$\texttt{combine}(\mathscr{T}) = \{C_1 \cap \cdots \cap C_n : C_1 \in T_1, \dots, C_n \in \mathscr{T}_n\}, \tag{7}$$

i.e., the combination of all primitive traits $T \in \mathscr{T}$.[7]

All this leads to a more explicit description: Given a priority ranking "$>$" on a set of traits $\mathscr{T}$, we can define a sequence of *nested* partitions: At the j-th level of the hierarchy we have the combination of j traits, while a path connecting the root node of Λ to said node represents the order in which these traits were introduced. For example, the instances of the composite trait "`type` and `brand`" (e.g., "Ford wagons" or "Toyota sedans") could be reached by first prescribing the car's `type` and then its `brand`, or the other way around. Either order reflects a certain priority so, more broadly, a priority ranking corresponds to the order in which traits are progressively combined to form composite ones. In more precise language, since $\texttt{combine}(\mathcal{S}) \preccurlyeq \texttt{combine}(\mathcal{S}')$ when $\mathcal{S}' \subseteq \mathcal{S}$ (see theorem 6 for a formal statement and proof), we conclude that

$$1 = \texttt{combine}(\mathcal{S}_0) \succcurlyeq \cdots \succcurlyeq \texttt{combine}(\mathcal{S}_j) \succcurlyeq \cdots \succcurlyeq \texttt{combine}(\mathcal{S}_n) = 0. \tag{8}$$

i.e., *a family of traits $\mathscr{T}$ on $\mathcal{A}$ equipped with a priority ranking respecting its logical structure gives rise to a nested sequence of partitions of $\mathcal{A}$ which also respects the logical structure of $\mathscr{T}$*. This reduction of a family of traits to a sequence of nested partitions—a *tree structure* on $\mathcal{A}$—is what forms the core of our reduction to the nested model.

Further properties of combinations of traits Recall here that the combination $\bigwedge_{T\in\mathcal{S}} T$ of a subset of traits $\mathcal{S} \subseteq \mathscr{T}$ is defined as the coarsest partition of $\mathcal{A}$ that refines all traits $T \in \mathcal{S}$. Under this definition, combining traits is easily seen to be commutative and associative; moreover, if we set

$$1 = \{\mathcal{A}\} \qquad \text{and} \quad 0 = \{\{a\} : a \in \mathcal{A}\}, \tag{9}$$

we also have

$$1 \wedge T = T \quad \text{and} \quad 0 \wedge T = 0. \tag{10}$$

for all $T \in \mathcal{A}$. Thus, from an algebraic viewpoint, combining traits behaves much like multiplication: the coarsest partition of $\mathcal{A}$ acts as an identity element for $\wedge$, while its finest partition acts as

7. A priori, the partition $\texttt{combine}(\mathscr{T})$ need not coincide with the finest possible partition of $\mathcal{A}$. For example, if the set of alternatives contains a yellow 2012 Chevy Impala and a yellow 2016 Impala, these two Impalas cannot be distinguished by either `type`, `brand`, `color`, or `model`; one would need to consider a new trait, `year`, to tell them apart. That said, if two alternatives are indistinguishable relative to *all* possible traits considered by an agent, we posit that they are, in fact, equivalent for said agent. By this token, we will be assuming in what follows that the set of primitive traits $\mathscr{T} = \{T_i\}_{i=1}^n$ provides a *complete description* of $\mathcal{A}$, i.e., $\texttt{combine}(\mathscr{T}) = \bigwedge_{T\in\mathscr{T}} T = 0$.

an absorbing element. Finally, it is also easy to check that "combine" preserves unions in the sense that

$$\texttt{combine}(\mathcal{S}_1 \cup \mathcal{S}_2) = \texttt{combine}(\mathcal{S}_1) \wedge \texttt{combine}(\mathcal{S}_2) \tag{11}$$

for all subsets $\mathcal{S}_1, \mathcal{S}_2$ of $\mathscr{T}$. In other words, combining traits satisfies the following properties for all $T, T', T'' \in \mathscr{T}$:

(a) *Commutativity:* $T \wedge T' = T' \wedge T$.

(b) *Associativity:* $T \wedge T' \wedge T'' = T \wedge (T' \wedge T'') = (T \wedge T') \wedge T''$.

(c) *Idempotency:* $T \wedge T = T$.

A key feature of the above is the interplay between primitives and conjunctions. Since a trait refines all its primitives ($T \preccurlyeq T'$ whenever $T \vdash T'$), we also have $T \wedge T' = T$ whenever $T \vdash T'$; in other words, the conjunction of a trait with any of its primitives returns the original trait unchanged. This suggests that the set of distinct composite traits that can be obtained from a set of "primitive" traits $\mathscr{T}$ is intimately linked to the prerequisite structure of $\mathscr{T}$. To formalize this, fix a logical structure $\vdash$ on $\mathscr{T} = \{T_i\}_{i=1}^n$, and write

$$\begin{aligned} \bigwedge^j(\mathscr{T}) &= \{\texttt{combine}(\{T_{i_1}, \cdots, T_{i_j}\}) : i_1, \ldots, i_j = 1, \ldots, n\} \\ &= \{T_{i_1} \wedge \cdots \wedge T_{i_j} : i_1, \ldots, i_j = 1, \ldots, n\} \end{aligned} \tag{12}$$

for the set of composite traits obtained by taking j-tuple conjunctions of elements of $\mathscr{T}$ ($j = 0, \ldots, n$). Subsequently, putting all these traits together, write

$$\bigwedge \equiv \bigwedge(\mathscr{T}) = \bigcup_{j=0}^{n} \bigwedge^j(\mathscr{T}) \tag{13}$$

for the collection of *all* composite traits generated by the primitive traits of $\mathscr{T}$. Of course, as we noted above, it is possible that different combinations of traits of $\mathscr{T}$ might lead to the *same* composite trait when primitives are involved (since adding primitives of already-included traits does not refine a composite trait).

In principle, the logical structure of $\mathscr{T}$ might preclude some element of $\bigwedge$—i.e., a conjunction of an *arbitrary* group of traits—from being obtained as the conjunction $\texttt{combine}(\mathcal{S})$ of the traits in a node $\mathcal{S} \in \Lambda$ of the decision lattice. In fact, our next result shows that this possibility is never realized: the set of composite traits $\bigwedge \equiv \bigwedge(\mathscr{T})$ obtained from primitive traits is naturally embedded in the associated similarity structure $\Lambda \equiv \Lambda(\mathscr{T}, \vdash)$.

Theorem 6 *Let $\mathscr{T} = \{T_i\}_{i=1}^n$ be a family of traits equipped with a logical structure "$\vdash$". Then, with notation as above, we have:*

(a) *Every element of $\bigwedge$ is of the form $\texttt{combine}(\mathcal{S})$ for some $\mathcal{S} \in \Lambda$; concretely,*

$$\bigwedge = \{\texttt{combine}(\mathcal{S}) : \mathcal{S} \in \Lambda\} \tag{14}$$

(b) *For all $\mathcal{S}, \mathcal{S}' \in \Lambda$, we have*

$$\texttt{combine}(\mathcal{S}) \preccurlyeq \texttt{combine}(\mathcal{S}') \quad \textit{whenever} \quad \mathcal{S} \supseteq \mathcal{S}'. \tag{15}$$

In words, the conjunction operator `combine` *is a surjective, order-preserving map – a lattice epimorphism – from* $(\Lambda, \supseteq)$ *to* $(\bigwedge, \preccurlyeq)$.

Proof To begin, we claim that $\texttt{prim}(T) \wedge T = T$ for all $T \in \mathscr{T}$: indeed, since $T \preccurlyeq T'$ for all $T' \in \texttt{prim}(\mathscr{T})$, this is a simple consequence of the associativity of $\wedge$ and the fact that $T \wedge T' = T$ whenever $T \preccurlyeq T'$.

With this auxiliary result at hand, let $T_{i_1} \wedge \cdots \wedge T_{i_j} \in \bigwedge$ be a composite trait generated by the primitive traits of $\mathscr{T}$. Then, letting $\mathcal{S}_j = \texttt{prim}(T_{i_j}) \cup \{T_{i_j}\}$ for all $j = 1, \dots, j$, we get

$$\begin{aligned} T_{i_1} \wedge \cdots \wedge T_{i_j} &= (\texttt{prim}(T_{i_1}) \wedge T_{i_1}) \wedge \cdots \wedge (\texttt{prim}(T_{i_j}) \wedge T_{i_j}) \\ &= \texttt{combine}(\texttt{prim}(T_{i_1}) \cup \{T_{i_1}\}) \wedge \cdots \wedge \texttt{combine}(\texttt{prim}(T_{i_j}) \cup \{T_{i_j}\}) \\ &= \texttt{combine}(\mathcal{S}_1) \wedge \cdots \wedge \texttt{combine}(\mathcal{S}_j) \\ &= \texttt{combine}(\mathcal{S}_1 \cup \cdots \cup \mathcal{S}_j), \end{aligned} \tag{16}$$

where the second and fourth lines follow directly from (11). Note also that we have $\texttt{prim}(T_{i_j}) \in \Lambda$ for all $j = 1, \dots, j$; moreover, since $T_{i_j} \in \texttt{prim}(T_{i_j})^+$ by definition, it follows that $\mathcal{S}_j \in \Lambda$ for all $j = 1, \dots, j$. Thus, with Λ closed under unions, we get $\mathcal{S} \equiv \mathcal{S}_1 \cup \cdots \cup \mathcal{S}_j$, and our claim follows.

Finally, for part (*b*), let $\mathcal{S}' = \{T_{i_1}, \dots, T_{i_{j'}}\} \subseteq \{T_{i_1}, \dots, T_{i_j}\} = \mathcal{S}$ for some $j' \leq j$. Then, by (11), we obtain

$$\begin{aligned} \texttt{combine}(\mathcal{S}) = T_{i_1} \wedge \ldots \wedge T_{i_j} &= (T_{i_1} \wedge \ldots \wedge T_{i_{j'}}) \wedge (T_{i_{j'+1}} \wedge \ldots \wedge T_{i_j}) \\ &\preccurlyeq T_{i_1} \wedge \ldots \wedge T_{i_{j'}} = \texttt{combine}(\mathcal{S}'), \end{aligned} \tag{17}$$

and our proof is complete. ■

This theorem serves as a sanity check that the nested presentation of a similarity structure is, indeed, faithful.

Appendix B. Proofs of Section 3

We provide here the proof of Proposition 1, which is restated below.

Proposition 1 (One-point lower bound under tree-induced similarity) *Let* $\rho \coloneqq \sigma_L$. *For* $K > 2$, $T > 32(K-1)\log(14)$, *and* $\rho > 0.22\sqrt{(K-1)/T}$, *any randomized one-point bandit algorithm satisfies*

$$\sup_{(y_t)_{t=1}^T \in \mathcal{Y}_\sigma^T} \mathbb{E}\big[\mathbf{Reg}_{T,1}(y_{1:T})\big] \;>\; \frac{1}{504}\sqrt{T(K-1)}.$$

Equivalently, for any such algorithm there exists a tree-compatible loss sequence $(y_t)_{t=1}^T \in \mathcal{Y}_\sigma^T$ *inducing* $\Omega(\sqrt{KT})$ *expected regret.*

Proof Let $\rho \coloneqq \sigma_L$. We define

$$C_\rho \coloneqq \Big\{x \in [0,1]^K : \max_{a,b \in \mathcal{A}} |x(a) - x(b)| \leq \rho\Big\} \qquad \text{and} \qquad C_\rho^T \coloneqq \Big\{(y_t)_{t=1}^T : \forall t,\ y_t \in C_\rho\Big\}.$$

Input: sequence of regularizers $(\psi_t)_{t\geq 1}$
Initialization: $p_1 = \arg\min_{p\in\Delta_K} \psi_1(p)$
for $t \geq 1$ **do**
  sample $A_t \sim p_t$ and set $z_t(a) = \sum_{j\in[L]} [\![a \sim_j A_t]\!] \frac{\ell_t(\mathrm{anc}_j(A_t))}{p_t[\mathrm{anc}_j(A_t)]}$ for all $a \in \mathcal{A}$
  set $p_{t+1} = \arg\min_{p\in\Delta_K} \langle \sum_{s\leq t} z_s, p\rangle + \psi_{t+1}(p)$
end

Algorithm 2: FTRL with Semi-Bandit Feedback

By Gerchinovitz and Lattimore (2016, Corollary 4), for $K > 2$, $T > 32(K-1)\log(14)$, and $\rho > 0.22\sqrt{(K-1)/T}$, any randomized one-point bandit algorithm satisfies

$$\sup_{(y_t)_{t=1}^T \in C_\rho^T} \mathbb{E}\big[\mathbf{Reg}_{T,1}(y_{1:T})\big] \;>\; \frac{1}{504}\sqrt{T(K-1)}.$$

It remains to show that $C_\rho^T \subseteq \mathcal{Y}_\sigma^T$. Fix any $x \in C_\rho$ and any $a, b \in \mathcal{A}$. Then $|x(a) - x(b)| \leq \rho = \sigma_L$. Since $s(a,b) \in [L]$ for $a \neq b$ and (σ_j) is nonincreasing, we have $\sigma_L \leq \sigma_{s(a,b)}$. Thus $|x(a) - x(b)| \leq \sigma_{s(a,b)}$, so x satisfies the tree-compatibility constraint. Therefore $C_\rho^T \subseteq \mathcal{Y}_\sigma^T$, and taking the supremum over $\mathcal{Y}_\sigma^T$ yields the proposition. ■

Appendix C. Proofs of Section 5

For convenience, we summarize the algorithm template used in Section 5 in Algorithm 2. We prove the following proposition before moving on to the proof of Theorem 2, which is restated afterwards.

Proposition 7 *Algorithm 2 with* $\psi_t(p) = \sqrt{5\max\{t, 20\}} \sum_{j\in[L]} R_j \sum_{v\in V_j} p[v] - \sqrt{p[v]}$ *satisfies that*

$$\mathbf{Reg}_{T,1} \leq 15 \sum_{j\in[L]} R_j \sqrt{|V_j|} + 2\sqrt{5} \sum_{t=1}^T \frac{1}{\sqrt{t}} \mathbb{E} \sum_{j\in[L]} R_j \sum_{v\in V_j} \sqrt{p_t[v]} - p_t[v]\,.$$

Proof This result is an application of Lemma 17 with $c = 5$; $\xi_j = 1$ for all $j \in [L]$; $M_t = [L]$ for all $t \geq 1$; and filtration $(\mathcal{F}_t)_{t\geq 0}$ with $\mathcal{F}_t$ being the σ-algebra generated by $(A_s)_{s\leq t}$ and $(\ell_s)_{s\leq t}$. In particular, as Lemma 17 requires, it holds that $\ell_t(v) \in [0, R_j]$ for all $j \in [L]$, $v \in V_j$, and $t \geq 1$; p_t is measurable w.r.t. $\mathcal{F}_{t-1}$; the law of A_t conditioned on $\mathcal{F}_{t-1}$ is p_t; and for all $a, a' \in \mathcal{A}$,

$$\sum_{v\succ a} \mathbb{E}_t[\ell_t(v) \mid v \succ A_t] - \sum_{w\succ a'} \mathbb{E}_t[\ell_t(w) \mid w \succ A_t] = \mathbb{E}_t[y_t(a) - y_t(a')]$$

simply since $\sum_{v\succ a} \ell_t(v) = y_t$ and $\ell_t(v)$ is independent of $[\![v \succ A_t]\!]$ conditioned on $\mathcal{F}_{t-1}$ as the adversary can only react to the learner's decisions that were made up to the previous round. ■

Theorem 2 *FTRL with* $\psi_t(p) = \sqrt{5\max\{t,20\}}\sum_{j\in[L]} R_j \sum_{v\in V_j} p[v] - \sqrt{p[v]}$ *and* $z_t(a) = \sum_{v\succ a}[\![v \succ A_t]\!]\ell_t(v)/p_t[v]$ *satisfies*

$$\mathbf{Reg}_{T,1} \leq \begin{cases} 15\sqrt{K_{\text{eff}}(R)} + 4\sqrt{5K_{\text{eff}}(R)T} \\ 30\sqrt{K_{\text{eff}}(R)} + 20K_{\text{eff},\Delta}(R)\ln(eT) \quad \textit{(under Condition 1)} \end{cases}$$

Proof Starting from the bound of Proposition 7, one can use the facts that $\sum_{v\in V_j}\sqrt{p_t[v]} - p_t[v] \leq \sqrt{|V_j|}$ for any $j \in [L]$ and that $\sum_{t=1}^T 1/\sqrt{t} \leq 2\sqrt{T}$ to directly obtain the first result. Concerning the second result, Lemma 14 implies that

$$2\sum_{j\in[L]} \frac{2\sqrt{5}R_j}{\sqrt{t}} \sum_{v\in V_j} \sqrt{p_t[v]} - p_t[v] \leq \langle \Delta, p_t\rangle + \frac{20}{t}\sum_{j\in[L]} R_j^2 \sum_{v\in V_j\colon \widetilde{\Delta}(v)\neq 0} \frac{1}{\widetilde{\Delta}(v)}$$

for any $\widetilde{\Delta}\colon \mathcal{V} \to \mathbb{R}$ such that (i) $\widetilde{\Delta}(v) \geq 0\ \forall v \in V$, (ii) $\Delta(a) \geq \sum_{v\succ a}\widetilde{\Delta}(v)\ \forall a \in \mathcal{A}$, and (iii) $|\{v \in V_j \mid \widetilde{\Delta}(v) = 0\}| \leq 1$ for all $j \in [L]$. Then, using that $\mathbb{E}\Delta(A_t) = \mathbb{E}\langle\Delta, p_t\rangle$, we obtain that

$$\mathbf{Reg}_{T,1} \leq 15\sum_{j\in[L]} R_j\sqrt{|V_j|} + \frac{1}{2}\mathbb{E}\left[\sum_{t=1}^T \Delta(A_t)\right] + 10\sum_{t=1}^T \frac{1}{t}\sum_{j\in[L]} R_j^2 \sum_{v\in V_j\colon \widetilde{\Delta}(v)\neq 0} \frac{1}{\widetilde{\Delta}(v)}\,.$$

Then, using that $\sum_{t=1}^T 1/t \leq \ln(eT)$ and that, by assumption, the second term is at most $\mathbf{Reg}_{T,1}/2$, we get that

$$\mathbf{Reg}_{T,1} \leq 30\sum_{j\in[L]} R_j\sqrt{|V_j|} + 20\sum_{j\in[L]} R_j^2 \sum_{v\in V_j\colon \widetilde{\Delta}(v)\neq 0} \frac{\ln(eT)}{\widetilde{\Delta}(v)}\,. \tag{18}$$

To reach a more concrete bound we impose now a simple form on $\widetilde{\Delta}$. In particular, let $\widehat{\Delta}(v) \coloneqq \min_{a\prec v}\Delta(a)$, and fixing any distribution $q \in \Delta_L$ (with strictly positive weights) over the levels of the tree, we take $\widetilde{\Delta}(v) = q(j)\widehat{\Delta}(v)$ for $v \in V_j$. This is a valid choice since it is non-negative, satisfies

$$\sum_{v\succ a}\widetilde{\Delta}(v) = \sum_{j\in[L]} q(j)\widehat{\Delta}(\text{anc}_j(a)) \leq \sum_{j\in[L]} q(j)\Delta(a) = \Delta(a)$$

for any $a \in \mathcal{A}$, and can only vanish at one node per level by the assumption that $|\{a \in \mathcal{A}\colon \Delta(a) = 0\}| \leq 1$. It then follows from (18) that

$$\mathbf{Reg}_{T,1} \leq 30\sum_{j\in[L]} R_j\sqrt{|V_j|} + 20\sum_{j\in[L]} \frac{R_j^2}{q(j)} \sum_{v\in V_j\colon \widehat{\Delta}(v)\neq 0} \frac{\ln(eT)}{\widehat{\Delta}(v)}\,.$$

The sought bound now follows by choosing $q(j) \propto R_j\sqrt{\sum_{v\in V_j\colon \widehat{\Delta}(v)\neq 0} 1/\widehat{\Delta}(v)}$.

■

Input: sequence of regularizers $(\psi_t)_{t\geq 1}$, sequence of constants $(b_j)_{j=1}^L$, number of extra queried points $n \in [L]$, distribution Ξ over subsets of $[L]$ of cardinality n
Define: for $j \in [L]$, let $\xi_j \coloneqq \sum_{S\in\mathrm{supp}(\Xi)} [\![j \in S]\!] \Xi(S)$
Initialization: $p_1 = \arg\min_{p\in\Delta_K} \psi_1(p)$

for $t \geq 1$ **do**
  sample $A_{t,1} \sim p_t$
  sample $M_t \sim \Xi$
  set $M_t^{\downarrow} = \{j \in M_t \mid j+1 \notin M_t\}$
  for $j \in M_t^{\downarrow}$ **do** set $\mathrm{smp}_t(j) = A_{t,1}$
  $i \leftarrow 2$
  for $j \in M_t$ **do**
    sample $A_{t,i} \sim p_t \mid A_{t,i} \sim_{j-1} A_{t,1}$
    set $\mathrm{smp}_t(j-1) = A_{t,i}$
    $i \leftarrow i+1$
  end
  for all $a \in \mathcal{A}$, set $z_t(a) = \sum_{j\in M_t} [\![a \sim_j A_{t,1}]\!] \dfrac{y_t(\mathrm{smp}_t(j)) - y_t(\mathrm{smp}_t(j-1)) + b_j}{\xi_j\, p_t[\mathrm{anc}_j(A_{t,1})]}$
  set $p_{t+1} = \arg\min_{p\in\Delta_K} \langle \sum_{s\leq t} z_s, p\rangle + \psi_{t+1}(p)$
end

Algorithm 3: FTRL with $n+1$-Point Bandit Feedback

Appendix D. Proofs of Section 6

We provide here a generalization of Algorithm 1 to handle $n+1$-point feedback, with n between 2 and L. We prove a generic regret bound for this algorithm, from which Theorem 3 (restated afterwards) follows as a corollary.

Theorem 8 *Algorithm 3 with $\psi_t \colon \mathbb{R}_+^d \to \mathbb{R}$ given by*

$$\psi_t(x) \coloneqq \frac{4\sqrt{c}}{\sqrt{(c-1)}} \sum_{j\in[L]} \frac{\sigma_j}{\sqrt{\xi_j}} \max\Big\{\sqrt{t}, \sqrt{c(c-1)/\xi_j}\Big\} \sum_{v\in V_j} x[v] - \sqrt{x[v]}$$

where $c \geq 2$, and $b_j = \sigma_j$ for all $j \in [L]$ satisfies

$$\mathbf{Reg}_{T,n+1} \leq 6c \sum_{j\in[L]} \frac{\sigma_j}{\xi_j} \sqrt{|V_j|} + \frac{16\sqrt{c}}{\sqrt{(c-1)}} \sum_{j\in[L]} \frac{\sigma_j}{\sqrt{\xi_j}} \sqrt{|V_j| T}\,.$$

Further, if Condition 1 holds, then, Algorithm 3 also satisfies

$$\mathbf{Reg}_{T,n+1} \leq 12c \sum_{j\in[L]} \frac{\sigma_j}{\xi_j} \sqrt{|V_j|} + \frac{64c}{c-1} \left(\sum_{j\in[L]} \frac{\sigma_j}{\sqrt{\xi_j}} \sqrt{\Gamma(j,\Delta)} \right)^2 \ln(eT)\,.$$

Proof Let $\mathcal{F}_t$ be the σ-algebra generated by $(A_{s,i})_{s\leq t, i\in[n+1]}$, $(M_s)_{s\leq t}$, and $(y_s)_{s\leq t}$. In the sequel, we will for brevity denote $\mathbb{P}(\cdot \mid \mathcal{F}_{t-1})$ by $\mathbb{P}_t(\cdot)$ and $\mathbb{E}[\cdot \mid \mathcal{F}_{t-1}]$ by $\mathbb{E}_t[\cdot]$. Clearly, $\mathbb{P}_t(A_{t,1} = a) =$

$p_t(a)$. For $2 \leq i \leq n+1$, we posit that a fixed function $\zeta_i \colon \operatorname{supp}(\Xi) \to \{0, \dots, L-1\}$ gives, at every round t, the unique level $\zeta_i(M_t)$ such that $A_{t,i} = \operatorname{smp}_t(\zeta_i(M_t))$. Notice then that for $2 \leq i \leq n+1$,

$$\begin{aligned}
\mathbb{P}_t(A_{t,i} = a) &= \sum_{a' \in \mathcal{A}, S \in \operatorname{supp}(\Xi)} \mathbb{P}_t\big(A_{t,i} = a \mid A_{t,1} = a', M_t = S\big) \mathbb{P}_t\big(A_{t,1} = a', M_t = S\big) \\
&= \sum_{a' \in \mathcal{A}, S \in \operatorname{supp}(\Xi)} \mathbb{P}_t\big(A_{t,i} = a \mid A_{t,1} = a', M_t = S\big) \mathbb{P}_t\big(A_{t,1} = a'\big) \mathbb{P}_t\big(M_t = S\big) \\
&= \sum_{a' \in \mathcal{A}, S \in \operatorname{supp}(\Xi)} [\![a' \sim_{\zeta_i(S)} a]\!] \frac{p_t(a)}{p_t[\operatorname{anc}_{\zeta_i(S)}(a')]} p_t(a') \Xi(S) \\
&= p_t(a) \sum_{S \in \operatorname{supp}(\Xi)} \Xi(S) \sum_{a' \in \mathcal{A}} [\![a' \sim_{\zeta_i(S)} a]\!] \frac{1}{p_t[\operatorname{anc}_{\zeta_i(S)}(a')]} p_t(a') \\
&= p_t(a) \sum_{S \in \operatorname{supp}(\Xi)} \Xi(S) \frac{1}{p_t[\operatorname{anc}_{\zeta_i(S)}(a)]} \sum_{a' \in \mathcal{A}} [\![a' \sim_{\zeta_i(S)} a]\!] p_t(a') = p_t(a) \,,
\end{aligned}$$

where the second equality uses that A_t and S_t are independent conditioned on $\mathcal{F}_{t-1}$, and the third equality follows from the sampling rules of Algorithm 3. Combining this with the fact that y_t is independent of the learner's actions at round t conditioned on $\mathcal{F}_{t-1}$, we get that

$$\mathbb{E}\left[\frac{1}{n+1} \sum_{i=1}^{n+1} y_t(A_{t,i})\right] = \mathbb{E}\big[y_t(A_{t,1})\big] \,.$$

Thus,

$$\mathbf{Reg}_{T,n+1} = \mathbb{E}\left[\sum_{t=1}^{T} y_t(A_{t,1})\right] - \min_{a \in \mathcal{A}} \mathbb{E}\left[\sum_{t=1}^{T} y_t(a)\right] . \tag{19}$$

Towards applying Lemma 17, fix a mapping $\lambda \colon \overline{\mathcal{V}} \to \mathcal{A}$ such that $v \succ \lambda(v)$ for all $v \in \overline{\mathcal{V}}$. Further, for each $v \in \overline{\mathcal{V}}$, let $(\omega_t(v))_{t \geq 1}$ be a random sequence of actions given by

$$\omega_t(v) = \begin{cases} \operatorname{smp}_t(\operatorname{lvl}(v)) & \text{if } v \succ A_{t,1} \text{ and } \{\operatorname{lvl}(v), \operatorname{lvl}(v)+1\} \cap M_t \neq \emptyset \\ \lambda(v) & \text{otherwise} \,. \end{cases}$$

Note that for $a \in \mathcal{A}$, $\omega_t(a) = a$. Next, we define at every round t a mapping $\ell_t \colon \mathcal{V} \to \mathbb{R}$ given by

$$\ell_t(v) \coloneqq y_t(\omega_t(v)) - y_t(\omega_t(\operatorname{prt}(v))) + \sigma_{\operatorname{lvl}(v)} \,.$$

Note that for any $v \in V_j$,

$$|\ell_t(v)| \leq 2\sigma_j \quad \text{and} \quad \ell_t(v) \geq 0 \,.$$

Further, define

$$\widetilde{\ell}_t(v) \coloneqq \mathbb{E}_t\big[\ell_t(v) \mid v \succ A_{t,1}, \operatorname{lvl}(v) \in M_t\big] \,,$$

so that

$$\begin{aligned}
\widetilde{\ell}_t(v) = \mathbb{E}_t\big[y_t(\omega_t(v)) \mid v \succ A_{t,1}, \operatorname{lvl}(v) \in M_t\big] \\
- \mathbb{E}_t\big[y_t(\omega_t(\operatorname{prt}(v))) \mid v \succ A_{t,1}, \operatorname{lvl}(v) \in M_t\big] + \sigma_{\operatorname{lvl}(v)} \,.
\end{aligned}$$

We aim to show now that the sum of the first two terms over $v \in \mathcal{V}\colon v \succ a$ is a telescoping sum for any action $a \in \mathcal{A}$. Towards that, we can rewrite the second term as follows for any $j \in [L]$ and $v \in V_j$:

$$\begin{aligned}\mathbb{E}_t\big[y_t(\omega_t(\mathrm{prt}(v))) \mid v \succ A_{t,1}, \mathrm{lvl}(v) \in M_t\big] &= \mathbb{E}_t\big[y_t(\omega_t(\mathrm{prt}(v))) \mid v = \mathrm{anc}_j(A_{t,1}), j \in M_t\big] \\ &= \mathbb{E}_t\big[y_t(\mathrm{smp}_t(j-1)) \mid v = \mathrm{anc}_j(A_{t,1}), j \in M_t\big] \\ &= \mathbb{E}_t\Big[\mathbb{E}_t\big[y_t(\mathrm{smp}_t(j-1)) \mid v = \mathrm{anc}_j(A_{t,1}), j \in M_t, y_t\big]\Big] \\ &= \mathbb{E}_t\Big[\textstyle\sum_{a\in\mathcal{A}} [\![\mathrm{prt}(v) \succ a]\!] \dfrac{p_t(a)}{p_t[\mathrm{prt}(v)]} y_t(a)\Big]\,,\end{aligned}$$

where the second equality follows from the definition of ω_t, the third is an application of the tower rule; and the fourth follows from the sampling rules of Algorithm 3, the fact that $\mathrm{prt}(v) = \mathrm{anc}_{j-1}(A_{t,1})$ whenever $v = \mathrm{anc}_j(A_{t,1})$, and the fact that y_t is drawn by the environment independently of the learner's actions at round t, conditioned on $\mathcal{F}_{t-1}$. At the same time, it similarly holds that

$$\begin{aligned}\mathbb{E}_t\big[y_t(\omega_t(v)) \mid v \succ A_{t,1}, \mathrm{lvl}(v) \in M_t\big] &= \mathbb{E}_t\big[y_t(\omega_t(v)) \mid v = \mathrm{anc}_j(A_{t,1}), j \in M_t\big] \\ &= \mathbb{E}_t\big[y_t(\mathrm{smp}_t(j)) \mid v = \mathrm{anc}_j(A_{t,1}), j \in M_t\big] \\ &= \mathbb{E}_t\Big[\mathbb{E}_t\big[y_t(\mathrm{smp}_t(j)) \mid v = \mathrm{anc}_j(A_{t,1}), j \in M_t, y_t\big]\Big] \\ &= \mathbb{E}_t\Big[\textstyle\sum_{a\in\mathcal{A}} [\![v \succ a]\!] \dfrac{p_t(a)}{p_t[v]} y_t(a)\Big]\,,\end{aligned}$$

where in the last step we used that, at round t, given that $\mathrm{anc}_j(A_{t,1})$ and $j \in M_t$, the conditional law (w.r.t. $\mathcal{F}_{t-1}$) of $\mathrm{smp}_t(j)$ is $[\![v \succ \cdot]\!] p_t(\cdot)/p_t[v]$ regardless of whether j belongs to $M_t^{\downarrow}$ or $M_t \setminus M_t^{\downarrow}$; in the latter case, this holds directly via the sampling rule of Algorithm 3 for $\mathrm{smp}_t(j)$, while in the former, we use that $\mathrm{smp}_t(j) = A_{t,1}$ and that conditioned on $\mathcal{F}_{t-1}$, the law of $A_{t,1}$ is p_t. We can then conclude that for any $a \in \mathcal{A}$,

$$\sum_{v\succ a} \widetilde{\ell}_t(v) = \mathbb{E}_t\big[y_t(a)\big] - \sum_{a'\in\mathcal{A}} p_t(a')\mathbb{E}_t\big[y_t(a')\big] + \sum_{j\in[L]} \sigma_j\,.$$

Implying that for any $a, a' \in \mathcal{A}$,

$$\sum_{v\succ a} \widetilde{\ell}_t(v) - \sum_{w\succ a'} \widetilde{\ell}_t(w) = \mathbb{E}_t\big[y_t(a) - y_t(a')\big]\,.$$

Notice, moreover, that

$$\begin{aligned}
&\sum_{v\succ a}[\![v\succ A_{t,1},\mathrm{lvl}(v)\in M_t]\!]\frac{\ell_t(v)}{\xi_{\mathrm{lvl}(v)}p_t[v]}\\
&\quad=\sum_{v\succ a}[\![v\succ A_{t,1},\mathrm{lvl}(v)\in M_t]\!]\frac{y_t(\omega_t(v))-y_t(\omega_t(\mathrm{prt}(v)))+\sigma_{\mathrm{lvl}(v)}}{\xi_{\mathrm{lvl}(v)}p_t[v]}\\
&\quad=\sum_{j\in[L]}[\![a\sim_j A_{t,1},j\in M_t]\!]\frac{y_t\big(\omega_t(\mathrm{anc}_j(a))\big)-y_t\big(\omega_t(\mathrm{anc}_{j-1}(a))\big)+\sigma_j}{\xi_j\,p_t[\mathrm{anc}_j(a)]}\\
&\quad=\sum_{j\in[L]}[\![a\sim_j A_{t,1},j\in M_t]\!]\frac{y_t\big(\omega_t(\mathrm{anc}_j(A_{t,1}))\big)-y_t\big(\omega_t(\mathrm{anc}_{j-1}(A_{t,1}))\big)+\sigma_j}{\xi_j\,p_t[\mathrm{anc}_j(A_{t,1})]}\\
&\quad=\sum_{j\in M_t}[\![a\sim_j A_{t,1}]\!]\frac{y_t\big(\omega_t(\mathrm{anc}_j(A_{t,1}))\big)-y_t\big(\omega_t(\mathrm{anc}_{j-1}(A_{t,1}))\big)+\sigma_j}{\xi_j\,p_t[\mathrm{anc}_j(A_{t,1})]}\\
&\quad=\sum_{j\in M_t}[\![a\sim_j A_{t,1}]\!]\frac{y_t(\mathrm{smp}_t(j))-y_t(\mathrm{smp}_t(j-1))+\sigma_j}{\xi_j\,p_t[\mathrm{anc}_j(A_{t,1})]}=z_t(a)\,,
\end{aligned}$$

where the last step follows form the definition of ω_t. Finally, it is easily verifiable that in our case, $\mathbb{P}_t(j\in M_t)=\xi_j\ \forall j\in[L]$, and $[\![v\succ A_t]\!]\perp\!\!\!\perp[\![\mathrm{lvl}(v)\in M_t]\!]\mid\mathcal{F}_{t-1}\ \forall v\in\mathcal{V}$. We can now invoke Lemma 17 with $R_j=2\sigma_j$ for $j\in[L]$; $A_t=A_{t,1}$ for $t\geq 1$; and ξ_j, M_t, y_t, ℓ_t, and $(\mathcal{F}_t)_{t\geq 0}$ as used in the current context; yielding that

$$\begin{aligned}
\mathbb{E}\left[\sum_{t=1}^{T}y_t(A_{t,1})\right]-\min_{a\in\mathcal{A}}\mathbb{E}\left[\sum_{t=1}^{T}y_t(a)\right]\leq\frac{8\sqrt{c}}{\sqrt{(c-1)}}\sum_{t=1}^{T}\frac{1}{\sqrt{t}}\sum_{j\in[L]}\frac{\sigma_j}{\sqrt{\xi_j}}\mathbb{E}\sum_{v\in V_j}\big(\sqrt{p_t[v]}-p_t[v]\big)\\
+6c\sum_{j\in[L]}\frac{\sigma_j}{\xi_j}\sqrt{|V_j|}\,.
\end{aligned}$$

which is also a bound on $\mathbf{Reg}_{T,n+1}$ via (19). At this junction, the rest of the proof proceeds similarly to the proof of Theorem 2. ■

Theorem 3 *Algorithm 1 with* $\psi_t(p)=2\sqrt{6}\sum_{j\in[L]}(\sigma_j/\sqrt{\delta_j})\sqrt{\max\{t,6/\delta_j\}}\sum_{v\in V_j}x[v]-\sqrt{x[v]}$ *and* $b_j=\sigma_j$ *for all* $j\in[L]$ *satisfies*

$$\mathbf{Reg}_{T,2}\leq\begin{cases}18\sqrt{K_{\mathrm{eff}}(\sigma/\delta)}+8\sqrt{6}\sqrt{K_{\mathrm{eff}}(\sigma/\sqrt{\delta})T}\\ 36\sqrt{K_{\mathrm{eff}}(\sigma/\delta)}+96K_{\mathrm{eff},\Delta}(\sigma/\sqrt{\delta})\ln(eT)\quad\textit{(under Condition 1)}\end{cases}.$$

Proof This follows as a special case of Theorem 8 choosing $c=3$ and $M_t=\{j_t\}$ at every round. ■

Appendix E. Proofs of Section 7

E.1. Proof of Corollary 4

We provide here the proof of Corollary 4, which is restated below.

Corollary 4 *Suppose we run Algorithm 1 on the tree $\mathcal{T}\big((\overline{\mathcal{A}}_j)_{j=0}^L\big)$ with parameters ψ_t, b_j as defined in Theorem 3, $\sigma_j = c2^{3-j}$, and $\delta_j \propto 2^{-2j/3}|\mathcal{A}_j|^{1/3}$. Then, assuming that $\min_{j\in[L]}\delta_j \geq 1/T$, it holds that*

$$\mathbf{Reg}_{T,2} \leq 16\cdot(9+4\sqrt{6})\sqrt{K_{\text{eff}}\big((\mathcal{A}_j)_{j=0}^L\big)T} + c\,2^{-L}T\,,$$

where $K_{\text{eff}}\big((\mathcal{A}_j)_{j=0}^L\big) \coloneqq c^2\left(\sum_{j\in[L]} 2^{-2j/3}|\mathcal{A}_j|^{1/3}\right)^3$.

Proof Firstly, since $\mathcal{A}_L$ is a $c2^{-L}$-cover of $\mathcal{A}$, we have that

$$\mathbf{Reg}_{T,2} \leq \mathbb{E}\left[\sum_{t=1}^T \frac{1}{2}\big(y_t(A_{t,1}) + y_t(A_{t,2})\big)\right] - \min_{a\in\mathcal{A}_L}\mathbb{E}\left[\sum_{t=1}^T y_t(a)\right] + c\,2^{-L}T\,.$$

So, using Theorem 3 gives that

$$\begin{aligned}\mathbf{Reg}_{T,2} &\leq 18\sqrt{K_{\text{eff}}(\sigma/\delta)} + 8\sqrt{6}\sqrt{K_{\text{eff}}(\sigma/\sqrt{\delta})T} + c\,2^{-L}T\\ &\leq 2(9+4\sqrt{6})\sqrt{K_{\text{eff}}(\sigma/\sqrt{\delta})T} + c\,2^{-L}T\,,\end{aligned}$$

where we have used in the second step that $\sqrt{K_{\text{eff}}(\sigma/\delta)} \leq \sqrt{K_{\text{eff}}(\sigma/\sqrt{\delta})T}$ by the assumption that $\min_{j\in[L]}\delta_j \geq 1/T$. Next, observe that

$$\sqrt{K_{\text{eff}}(\sigma/\sqrt{\delta})} = \sum_{j\in[L]}\frac{\sigma_j}{\sqrt{\delta_j}}\sqrt{|\overline{\mathcal{A}}_j|} = c2^3\sum_{j\in[L]}\frac{2^{-j}}{\sqrt{\delta_j}}\sqrt{|\overline{\mathcal{A}}_j|} \leq c2^3\sum_{j\in[L]}\frac{2^{-j}}{\sqrt{\delta_j}}\sqrt{|\mathcal{A}_j|}\,,$$

where the last step uses that $\overline{\mathcal{A}}_j \subseteq \mathcal{A}_j$. Finally, our choice for δ_j entails that

$$\sum_{j\in[L]}\frac{2^{-j}}{\sqrt{\delta_j}}\sqrt{|\mathcal{A}_j|} = \left(\sum_{j\in[L]} 2^{-2j/3}|\mathcal{A}_j|^{1/3}\right)^{3/2}\,.$$

■

E.2. On the Near-Optimality of the Tree Constructed in Section 7

We will assume here that $\mathcal{A}$ is finite (with cardinality K) and that T is large enough so that $d_{\min} \coloneqq \min_{a,b\in\mathcal{A}:\, a\neq b} d_{\mathcal{A},\infty}(a,b) \geq \sqrt{K/T}$. The effect of this is that executing Algorithm 1 on a strict subset of the action set would force a worst-case approximation error of order $\sqrt{KT}$, which would nullify any attempt to leverage the structure. Now, any nested sequence $(\mathcal{P}_j)_{j=0}^{L'}$ of partitions of $\mathcal{A}$ such that $\mathcal{P}_0 = \{\mathcal{A}\}$ and $\mathcal{P}_{L'} = \{\{a\}\}_{a\in\mathcal{A}}$ can be used to build a tree of L' levels with its leaves

coinciding with the action set. In a similar manner to Corollary 4, executing Algorithm 1 on this tree results in a regret bound featuring a K_{eff} of (with the optimal choice of δ)

$$\left(\sum_{j=1}^{L'} \sigma_{*,j}^{2/3} |\mathcal{P}_j|^{1/3}\right)^3 \quad \text{where } \sigma_{*,j} \coloneqq \max_{S \in \mathcal{P}_{j-1}} \max_{a,b \in S} d_{\mathcal{A},\infty}(a,b)\,.$$

It is not clear if directly optimizing this quantity in terms of the nested structure is a feasible endeavor. Instead, we show in the following lemma that the recipe we described for constructing the tree—by forming a $c2^{-j}$ cover of $\mathcal{A}$ at level j—minimizes this quantity up to small constants, provided that these covers are optimal. Fortunately, near-optimal covers can be constructed in time polynomial in K. At any level j, we can associate each action $a \in \mathcal{A}$ with the set $\{a' \in \mathcal{A} \colon d_{\mathcal{A},\infty}(a,a') \leq c2^{-j}\}$; so, building an optimal $c2^{-j}$-cover reduces to the set cover problem, which can be solved in polynomial time using a simple greedy algorithm (Johnson, 1974). In particular, using this greedy algorithm, one builds a cover that is at worst an extra factor of $1 + \ln K$ larger than the minimum size (Johnson, 1974; Slavík, 1997).

Lemma 9 *Set $c = \max_{a,b \in \mathcal{A}} d_{\mathcal{A},\infty}(a,b)$ and $L = 1 + \lfloor -\log_2(c^{-1} d_{\min}) \rfloor$. Let $\mathcal{A}_0 \coloneqq \{a_0\}$ for some arbitrary action a_0 and $(\mathcal{A}_j)_{j \in [L]}$ be a sequence of subsets of $\mathcal{A}$ such that $\mathcal{A}_j$ is an optimal $c2^{-j}$-cover of $\mathcal{A}$. Then, for any nested sequence $(\mathcal{P}_i)_{i=0}^{L'}$ of partitions of $\mathcal{A}$ such that $\mathcal{P}_0 = \{\mathcal{A}\}$ and $\mathcal{P}_{L'} = \{\{a\}\}_{a \in \mathcal{A}}$, it holds that*

$$c^2 \left(\sum_{j=1}^{L} 2^{-2j/3} |\mathcal{A}_j|^{1/3}\right)^3 \leq 20 \left(\sum_{i=1}^{L'} \sigma_{*,i}^{2/3} |\mathcal{P}_i|^{1/3}\right)^3,$$

where $\sigma_{*,i} \coloneqq \max_{S \in \mathcal{P}_{i-1}} \max_{a,b \in S} d_{\mathcal{A},\infty}(a,b)$.

Note that this choice of L ensures that $c2^{-L} < d_{\min}$, which implies that $\mathcal{A}_L = \mathcal{A}$.

Proof We begin by defining a mapping $f \colon [L] \to [L']$ as follows:

$$f(j) \coloneqq \min\Big\{ i \in [L'] \mid \max_{a,b \in S} d_{\mathcal{A},\infty}(a,b) < c2^{-j} \;\forall S \in \mathcal{P}_i \Big\}\,. \tag{20}$$

Since $\mathcal{P}_{L'} = \{\{a\}\}_{a \in \mathcal{A}}$, $\max_{a,b \in S} d_{\mathcal{A},\infty}(a,b) = 0$ for all $S \in \mathcal{P}_{L'}$; thus, the set in the R.H.S. of (20) is never empty. Note that the pre-image of $i \in [L']$, if non-empty, is an interval $B_i \subseteq [L]$ satisfying that

$$\max_{j \in B_i} c2^{-j} \leq \sigma_{*,i}\,.$$

This is easy to see for $i = 1$: since we set $c = \max_{a,b \in \mathcal{A}} d_{\mathcal{A},\infty}(a,b)$, we have that $\sigma_{*,1} = c$. (Recall that $\sigma_{*,i}$ concerns the maximum diameter of the subsets of $\mathcal{P}_{i-1}$, not $\mathcal{P}_i$.) For $i \geq 2$, if $c2^{-j} > \sigma_{*,i}$ for a certain $j \in B_i$, then it must be that $j \in B_{i'}$ for some $i' \in \{1, \dots, i-1\}$, which is a contradiction. It then follows that

$$\sum_{j \in B_i} \left(c2^{-j}\right)^{2/3} \leq \sigma_{*,i}^{2/3} \sum_{r=0}^{\infty} 2^{-2r/3} \leq \frac{2^{2/3}}{2^{2/3} - 1} \sigma_{*,i}^{2/3}\,.$$

This allows concluding that

$$c^2\left(\sum_{j=1}^{L} 2^{-2j/3}|\mathcal{P}_{f(j)}|^{1/3}\right)^3 \leq 20\left(\sum_{i=1}^{L'}\sigma_{*,i}^{2/3}|\mathcal{P}_i|^{1/3}\right)^3 .$$

To obtain the sought result, it remains to use that $|\mathcal{P}_{f(j)}| \geq |\mathcal{A}_j|$. This holds since $\mathcal{A}_j$ is a $c2^{-j}$-cover of $\mathcal{A}$ with minimum cardinality, meantime, collecting an arbitrary member of each subset in $\mathcal{P}_{f(j)}$ results too in a $c2^{-j}$-cover of $\mathcal{A}$. ■

E.3. Proof of Corollary 5

We provide here the proof of Corollary 5, which is restated below.

Corollary 5 *Let $T \geq 8$. Algorithm 1 run on the tree $\mathcal{T}\big((\overline{\mathcal{A}}_j)_{j=0}^{L}\big)$ induced by the covers described above with $L = \lfloor \frac{1}{d}\log_2(T)\rfloor$ and parameters ψ_t, b_j as defined in Theorem 3, $\sigma_j = GD\sqrt{d}\,2^{3-j}$ and $\delta_j \propto 2^{j(d-2)/3}$, satisfies the following regret upper bound*

$$\mathbf{Reg}_{T,2} \lesssim GD \cdot \begin{cases} \sqrt{T} & \text{if } d = 1 \\ (\ln T)^{3/2}\sqrt{T} & \text{if } d = 2 \\ \sqrt{d}T^{\frac{d-1}{d}} & \text{if } d > 2 \end{cases} ,$$

where $\lesssim$ denotes an inequality up to a universal constant.

Proof Since $|\mathcal{A}_j| \leq 2^{jd}$ and $L = \lfloor \frac{1}{d}\log_2(T)\rfloor$, Corollary 4 gives that (supposing for now that $\min_j \delta_j \geq 1/T$)

$$\mathbf{Reg}_{T,2} \leq 16\cdot(9+4\sqrt{6})\cdot GD\sqrt{d}\cdot\left(\sum_{j\in[L]} 2^{j(d-2)/3}\right)^{3/2}\sqrt{T} + 2GD\sqrt{d}\,T^{\frac{d-1}{d}} .$$

Then, observe that

$$\left(\sum_{j\in[L]} 2^{j(d-2)/3}\right)^{3/2} \leq \begin{cases} L^{3/2} & \text{if } d = 2 \\ 11\cdot 2^{L(\frac{d}{2}-1)_+} & \text{if } d \neq 2 \end{cases}$$

where $(\,\cdot\,)_+ = \max\{\cdot, 0\}$. This leads to the stated regret bound after using that $L = \lfloor \frac{1}{d}\log_2(T)\rfloor$.

To conclude the proof, it only remains to check that $\max_{j\in[L]}\delta_j^{-1/2} \leq \sqrt{T}$ for $T \geq 8$. Indeed,

- if $d = 1$,

$$\delta_j^{-1/2} \leq \delta_L^{-1/2} = 2^{\frac{L}{6}}\left(\sum_{j=1}^{L} 2^{-\frac{j}{3}}\right)^{1/2} \leq 2\cdot 2^{\frac{L}{6}} \leq 2T^{\frac{1}{6}} \leq \sqrt{T}$$

  as soon as $T \geq 8$;
- if $d = 2$, $\delta_j^{-1/2} = \sqrt{L} \leq \sqrt{\log_2(T)/2} \leq \sqrt{T}$;

- if $d \geq 3$,

$$\delta_j^{-1/2} \leq \delta_1^{-1/2} = 2^{-\frac{d-2}{6}} \left(\sum_{j=1}^{L} 2^{j\left(\frac{d-2}{3}\right)} \right)^{1/2} \leq 2^{1+L\left(\frac{d-2}{6}\right)} \leq 2T^{\frac{d-2}{6d}} \leq 2T^{\frac{1}{6}} \leq \sqrt{T}$$

as soon as $T \geq 8$.

■

Appendix F. FTRL Analysis

F.1. Helper Lemmas

Lemma 10 *For $t \geq 1$, let $(\psi_t)_t$ be a sequence of functions where $\psi_t \colon \mathbb{R}^K \to (-\infty, +\infty]$. Additionally, let $(z_t)_t$ be an arbitrary sequence of vectors in $\mathbb{R}^K$. Define $p_1 \coloneqq \arg\min_{p \in \Delta_K} \psi_1(p)$ and for $t \geq 2$,*

$$p_t \coloneqq \arg\min_{p \in \Delta_K} \left\langle \textstyle\sum_{s \leq t-1} z_s, p \right\rangle + \psi_t(p) \,.$$

Then, assuming that the sequence $(p_t)_t$ is well-defined, it holds for any comparator $q \in \Delta_K$ and horizon $T \geq 1$ that

$$\sum_{t=1}^{T} \langle z_t, p_t - q \rangle \leq \psi_{T+1}(q) - \psi_1(p_1) + \sum_{t=1}^{T} \psi_t(p_{t+1}) - \psi_{t+1}(p_{t+1}) + \sum_{t=1}^{T} F_t\big(p_t, \textstyle\sum_{s \leq t} z_s\big) - F_t\big(p_{t+1}, \textstyle\sum_{s \leq t} z_s\big) \,,$$

where for $t \geq 1$, $z \in \mathbb{R}^d$, and $p \in \Delta_K$; we define $F_t(p, z) \coloneqq \langle z, p \rangle + \psi_t(p)$.

Proof For any $t \in [T]$, $p_t = \arg\min_{p \in \Delta_K} F_t\big(p, \sum_{s \leq t-1} z_s\big)$. So,

$$\begin{aligned} \sum_{t=1}^{T} \langle z_t, p_t - q \rangle &= \sum_{t=1}^{T} \langle z_t, p_t \rangle + \psi_{T+1}(q) - F_{T+1}(q, \textstyle\sum_{s \leq T} z_s) \\ &\leq \sum_{t=1}^{T} \langle z_t, p_t \rangle + \psi_{T+1}(q) - F_{T+1}(p_{T+1}, \textstyle\sum_{s \leq T} z_s) \,. \end{aligned}$$

Moving on,

$$\begin{aligned} -F_{T+1}(p_{T+1}, \textstyle\sum_{s \leq t} z_s) &= -\psi_1(p_1) + \psi_1(p_1) - F_{T+1}(p_{T+1}, \textstyle\sum_{s \leq T} z_s) \\ &= -\psi_1(p_1) + F_1(p_1, \mathbf{0}) - F_{T+1}(p_{T+1}, \textstyle\sum_{s \leq T} z_s) \\ &= -\psi_1(p_1) + \sum_{t=1}^{T} F_t(p_t, \textstyle\sum_{s \leq t-1} z_s) - F_{t+1}(p_{t+1}, \textstyle\sum_{s \leq t} z_s) \,. \end{aligned}$$

Lastly at any round t,

$$\begin{aligned}
&\langle z_t, p_t\rangle + F_t(p_t, \textstyle\sum_{s\le t-1} z_s) - F_{t+1}(p_{t+1}, \textstyle\sum_{s\le t} z_s) \\
&\quad = F_t(p_t, \textstyle\sum_{s\le t} z_s) - F_{t+1}(p_{t+1}, \textstyle\sum_{s\le t} z_s) \\
&\quad = F_t(p_t, \textstyle\sum_{s\le t} z_s) - F_t(p_{t+1}, \textstyle\sum_{s\le t} z_s) + \psi_t(p_{t+1}) - \psi_{t+1}(p_{t+1})\,,
\end{aligned}$$

concluding the proof. ■

We say that a convex function $f\colon \mathbb{R} \to (-\infty, +\infty]$ is Legendre if it is proper, closed, essentially smooth (see (Rockafellar, 1970, Chapter 26) for a definition) and strictly convex on $\mathrm{int}(\mathrm{dom}(f))$.

Lemma 11 *Let $f\colon \mathbb{R} \to (-\infty, +\infty]$ be a convex function of the Legendre type satisfying $\mathbb{R}_{++} \subseteq \mathrm{dom}(f)$. For any $p \in \Delta_K$, define*

$$\psi(p) \coloneqq \sum_{j\in[L]} \mu_j \sum_{v\in V_j} f(p[v])\,,$$

where $\mu_j > 0$ is a fixed weight for level j. Fix two vectors $y, z \in \mathbb{R}^K$, and let

$$q \coloneqq \operatorname*{arg\,min}_{p\in\Delta_K} \langle p, y\rangle + \psi(p)\,.$$

Then, for any $(c_j)_{j\in[L]} \in \mathbb{R}^L$ and $\ell\colon \mathcal{V} \to \mathbb{R}$ such that $z(a) = \sum_{v\succ a} \ell(v)$, it holds that

$$\begin{aligned}
\langle q, y+z\rangle + \psi(q) - \min_{p\in\Delta_K} \big(\langle p, y+z\rangle + \psi(p)\big) \\
\le \sum_{j\in[L]} \mu_j \sum_{v\in V_j} D_{f^*}\big(f'(p[v]) - (\ell(v) - c_j)/\mu_j \big\| f'(p[v])\big)\,.
\end{aligned}$$

Proof Associate to every $v \in \mathcal{V}$ a vector $h_v \in \mathbb{R}^K$ such that $h_v(a) \coloneqq [\![v \succ a]\!]$; hence, for any vector $x \in \mathbb{R}^K$,

$$x^\top h_v = \sum_{a\in\mathcal{A}\colon\, v\succ a} x(a) = x[v]\,.$$

Moreover, let $\eta_v \coloneqq 1/\mu_{\mathrm{lvl}(v)}$. So, for any $p \in \Delta_K$,

$$\psi(p) \coloneqq \sum_{v\in\mathcal{V}} \frac{1}{\eta_v} f(p^\top h_v)\,.$$

Hence, the result would follow from Proposition 26 if its conditions on f and $\{h_v\}_{v\in\mathcal{V}}$ hold here. By the assumption that f is Legendre, it is indeed proper, closed, and essentially smooth as required. Next, note that for any $v \in \mathcal{V}$, the definition of h_v together with the assumption that $\mathbb{R}_{++} \subseteq \mathrm{dom}(f)$ guarantee that $\mathbb{R}^K_{++} \subseteq \{x \in \mathbb{R}^K\colon h_v^\top x \in \mathrm{dom}(f)\}$. Hence,

$$\mathrm{relint}(\{x \in \mathbb{R}^K\colon h_v^\top x \in \mathrm{dom}(f)\}) = \mathrm{int}(\{x \in \mathbb{R}^K\colon h_v^\top x \in \mathrm{dom}(f)\}) \supseteq \mathbb{R}^K_{++}\,.$$

And since $\mathrm{relint}(\Delta_K) \subset \mathbb{R}^K_{++}$ and $\mathrm{relint}(\Delta_K) \neq \emptyset$, it holds, as required, that

$$\mathrm{relint}(\Delta_K) \cap \big(\cap_{v\in\mathcal{V}} \mathrm{relint}(\{x \in \mathbb{R}^K\colon h_v^\top x \in \mathrm{dom}(f)\})\big) \neq \emptyset\,.$$

■

Lemma 12 *Let $f\colon \mathbb{R}_+ \to (-\infty, +\infty]$ be a convex function of the Legendre type satisfying $\mathbb{R}_{++} \subseteq \operatorname{dom}(f)$. For any $p \in \Delta_K$, define*

$$\psi(p) \coloneqq \sum_{j\in[L]} \mu_j \sum_{v\in V_j} f(p[v])\,,$$

where $\mu_j > 0$ is a fixed weight for level j. It holds for any $p \in \Delta_K$ and $\zeta\colon \mathcal{V} \to \mathbb{R}$ that

$$-\psi(p) \leq \sum_{a\in\mathcal{A}} p(a) \sum_{v\succ a} \zeta(v) + \sum_{j\in[L]} \mu_j \sum_{v\in V_j} f^*(-\zeta(v)/\mu_j)\,.$$

Proof Using the same arguments in the proof of Lemma 11 (note that $\mu_j f$ is Legendre when f is Legendre), we can leverage the first result of Lemma 25 (taking $u = \mathbf{0}$) to obtain that

$$\psi(p) + \sum_{a\in\mathcal{A}} p(a) \sum_{v\succ a} \zeta(v) \geq -\sum_{j\in[L]} \mu_j \sum_{v\in V_j} f^*(-\zeta(v)/\mu_j)\,,$$

where we have used that $(\mu_j f)^*(\cdot) = \mu_j f^*(\cdot/\mu_j)$. ∎

F.1.1. SPECIFIC RESULTS FOR THE NESTED TSALLIS REGULARIZER

Lemma 13 *Let $f(x) = \frac{2}{\eta}(x - \sqrt{x})$ with $\eta > 0$. Then, for any $p > 0$, $c > 1$, and $y \in \mathbb{R}$, such that $y \geq -\frac{1}{c\eta\sqrt{p}}$, it holds that*

$$D_{f^*}\big(f'(p) - y \big\| f'(p)\big) \leq \frac{c}{c-1}\eta p^{3/2} y^2\,.$$

Proof We have that $f'(x) = \frac{2}{\eta}(1 - \frac{1}{2\sqrt{x}})$ and $f^*(z) = \frac{1}{\eta}\cdot\frac{1}{2-\eta z}$ for $\eta z < 2$. Note that

$$\eta f'(p) = 2 - \frac{1}{\sqrt{p}} < 2 \text{ and } \eta f'(p) - \eta y = 2 - \frac{1}{\sqrt{p}} - \eta y \leq 2 - \frac{1}{\sqrt{p}} + \frac{1}{c\sqrt{p}} < 2\,,$$

where we have used that $p > 0$ and that $\eta y \geq -\frac{1}{c\sqrt{p}}$ with $c > 1$. Hence,

$$\begin{aligned}
D_{f^*}\big(f'(p) - y \big\| f'(p)\big) &= yp + f^*(f'(p) - y) - f^*(f'(p)) \\
&= yp + \frac{1}{\eta}\sqrt{p}\left(\frac{1}{1+\eta y\sqrt{p}} - 1\right) \\
&= yp - \frac{yp}{1+\eta y\sqrt{p}} \\
&= \frac{\eta p^{3/2} y^2}{1+\eta y\sqrt{p}} \\
&\leq \frac{c}{c-1}\eta p^{3/2} y^2\,,
\end{aligned}$$

where the last inequality uses the assumption that $\eta y \geq -\frac{1}{c\sqrt{p}}$. ∎

Lemma 14 *Fix any $p \in \Delta_K$, $\mu_j > 0$ for $j \in [L]$, $\lambda \in \mathbb{R}$, $g \in \mathbb{R}^K$, and $\zeta \colon \mathcal{V} \to \mathbb{R}$ such that $g(a) + \lambda \geq \sum_{v \succ a} \zeta(v)$ for $a \in \mathcal{A}$ and $\zeta(v)/\mu_j > -2$ for $j \in [L]$ and $v \in V_j$. Then, it holds that*

$$2 \sum_{j \in [L]} \mu_j \sum_{v \in V_j} \sqrt{p_t[v]} - p_t[v] \leq \langle g, p \rangle + \lambda + \sum_{j \in [L]} \mu_j \sum_{v \in V_j} \frac{1}{2 + \zeta(v)/\mu_j} \,.$$

In particular, for any $\widetilde{\zeta} \colon \mathcal{V} \to \mathbb{R}$ such that (i) $\widetilde{\zeta}(v) \geq 0 \; \forall v \in V$, (ii) $g(a) \geq \sum_{v \succ a} \widetilde{\zeta}(v)$, and (iii) $|\{v \in V_j \mid \widetilde{\zeta}(v) = 0\}| \leq 1$ for all $j \in [L]$; it holds that

$$2 \sum_{j \in [L]} \mu_j \sum_{v \in V_j} \sqrt{p_t[v]} - p_t[v] \leq \langle g, p \rangle + \sum_{j \in [L]} \mu_j^2 \sum_{v \in V_j \colon \widetilde{\zeta}(v) \neq 0} \frac{1}{\widetilde{\zeta}(v)} \,.$$

Proof Let $f(x) = 2(x - \sqrt{x})$. Then $f^*(y) = 1/(2 - y)$ for $y < 2$ and $+\infty$ elsewhere. Hence, applying Lemma 12 with this function gives that

$$\begin{aligned} 2 \sum_{j \in [L]} \mu_j \sum_{v \in V_j} \sqrt{p_t[v]} - p_t[v] &\leq \sum_{a \in \mathcal{A}} p(a) \sum_{v \succ a} \zeta(v) + \sum_{j \in [L]} \mu_j \sum_{v \in V_j} \frac{1}{2 + \zeta(v)/\mu_j} \\ &\leq \langle g, p \rangle + \lambda + \sum_{j \in [L]} \mu_j \sum_{v \in V_j} \frac{1}{2 + \zeta(v)/\mu_j} \,, \end{aligned}$$

where the condition $\zeta(v)/\mu_j > -2$ for $v \in V_j$ ensures we respect the effective domain of f^*. Now, for the second part, choose $\lambda = -\sum_{j \in [L]} \mu_j$ and $\zeta(v) = \widetilde{\zeta}(v) - \mu_j$ for $v \in V_j$, getting as a consequence that

$$2 \sum_{j \in [L]} \mu_j \sum_{v \in V_j} \sqrt{p_t[v]} - p_t[v] \leq \langle g, p \rangle + \sum_{j \in [L]} \mu_j \left(\sum_{v \in V_j} \frac{1}{1 + \widetilde{\zeta}(v)/\mu_j} - 1 \right) .$$

Finally, to obtain the sought result we use that $1/(1 + \widetilde{\zeta}(v)/\mu_j) \geq \mu_j/\widetilde{\zeta}(v)$ for any v such that $\widetilde{\zeta}(v) \neq 0$ and that $1/(1 + \widetilde{\zeta}(v)/\mu_j) = 1$ otherwise, which we assumed to be the case for at most one node per level. ■

Lemma 15 *Fix $p \in \Delta_d$ and $x \in \mathbb{R}^d$. It holds that for any $j \in [d]$ that*

$$\sum_{i \in [d]} p(i)^{3/2} \big([\![i = j]\!] x(i)/p(i) - x(j) \big)^2 = x(j)^2 \left(p(j)^{-1/2} - 2p(j)^{1/2} + \textstyle\sum_{i \in [d]} p(i)^{3/2} \right) .$$

It also holds that

$$\sum_{j \in [d]} p(j) \left(p(j)^{-1/2} - 2p(j)^{1/2} + \textstyle\sum_{i \in [d]} p(i)^{3/2} \right) \leq 2 \sum_{j \in [d]} p(j)^{1/2} - p(j) \,.$$

Proof We have that

$$\begin{aligned}
&\sum_{i\in[d]} p(i)^{3/2}\big([\![i=j]\!]x(i)/p(i)-x(j)\big)^2\\
&\quad=\sum_{i\in[d]} p(i)^{3/2}\big([\![i=j]\!]x(j)/p(j)-x(j)\big)^2\\
&\quad=\sum_{i\in[d]} p(i)^{3/2}\big([\![i=j]\!]x(j)^2/p(j)^2+x(j)^2-2[\![i=j]\!]x(j)^2/p(j)\big)\\
&\quad=x(j)^2\left(p(j)^{-1/2}-2p(j)^{1/2}+\textstyle\sum_{i\in[d]}p(i)^{3/2}\right).
\end{aligned}$$

Next,

$$\begin{aligned}
&\sum_{j\in[d]} p(j)\left(p(j)^{-1/2}-2p(j)^{1/2}+\textstyle\sum_{i\in[d]}p(i)^{3/2}\right)\\
&\quad=\sum_{j\in[d]} p(j)^{1/2}-2\sum_{j\in[d]} p(j)^{3/2}+\sum_{j\in[d]} p(j)^{3/2}\\
&\quad=\sum_{j\in[d]} p(j)^{1/2}-\sum_{j\in[d]} p(j)^{3/2}\\
&\quad=\sum_{j\in[d]} p(j)^{1/2}(1-p(j))\\
&\quad=\sum_{j\in[d]} p(j)^{1/2}(1-p(j)^{1/2})(1+p(j)^{1/2})\\
&\quad\le 2\sum_{j\in[d]} p(j)^{1/2}(1-p(j)^{1/2})=2\sum_{j\in[d]} p(j)^{1/2}-p(j)\,.
\end{aligned}$$

■

F.2. Generic Regret Guarantees for the Nested Tsallis Regularizer

Lemma 16 *Let $(\eta_{t,j})_{t\ge1,j\in[L]}$ be a sequence of positive numbers. Accordingly, for every $t\ge1$, define $\psi_t\colon\mathbb{R}^k\to(-\infty,+\infty]$ as*

$$\psi_t(x)\coloneqq 2\sum_{j\in[L]}\frac{1}{\eta_{t,j}}\sum_{v\in V_j}\big(x[v]-\sqrt{x[v]}\big)$$

for $x\in\mathbb{R}^K$. Let $(z_t)_t$ be a sequence of loss vectors in $\mathbb{R}^K$, and define

$$p_1\coloneqq\operatorname*{arg\,min}_{p\in\Delta_K}\psi_1(p)\quad\text{and for }\ t\ge2,\ \ p_t\coloneqq\operatorname*{arg\,min}_{p\in\Delta_K}\big\langle\textstyle\sum_{s\le t-1}z_s,p\big\rangle+\psi_t(p)\,.$$

Fix $c>1$, and for each $t\ge1$, fix a sequence $(c_{t,j})_{j\in[L]}\in\mathbb{R}^L$ and a function $\ell_t\colon\mathcal{V}\to\mathbb{R}$ such that

$$z_t(a)=\sum_{v\succ a}\ell_t(v)\ \ \forall a\in\mathcal{A}\quad\text{and}\quad\ell_t(v)-c_{t,j}\ge-\frac{1}{c\eta_{t,j}\sqrt{p_t[v]}}\ \ \forall j\in[L]\,,\ v\in V_j\,.$$

Then, it holds for horizon $T \geq 1$ and comparator $q \in \Delta_k$ that

$$\sum_{t=1}^{T} \langle z_t, p_t - q \rangle \leq 2 \sum_{t=1}^{T+1} \sum_{j \in [L]} \widetilde{\eta}_{t,j} \sum_{v \in V_j} \left(\sqrt{p_t[v]} - p_t[v] \right) + \frac{c}{c-1} \sum_{t=1}^{T} \sum_{j \in [L]} \eta_{t,j} \sum_{v \in V_j} \left(p_t[v] \right)^{3/2} (\ell_t(v) - c_{t,j})^2 \,,$$

where $\widetilde{\eta}_{1,j} = \eta_{1,j}^{-1}$ and for $t \geq 2$, $\widetilde{\eta}_{t,j} = \eta_{t,j}^{-1} - \eta_{t-1,j}^{-1}$.

Proof This is obtained by combining Lemmas 10, 11 and 13, and using the fact that $\psi_{T+1}(q) \leq 0$. ■

Lemma 17 *Fix a sequence of positive numbers $(R_j)_{j \in [L]}$ and a sequence $(\xi_j)_{j \in [L]}$ of weights in $[0,1]$. Moreover, with $t \geq 1$, let $(y_t)_t$ be a random sequence of loss functions each mapping $\mathcal{A}$ to $\mathbb{R}$,$(A_t)_t$ a random sequence of actions from $\mathcal{A}$, $(M_t)_t$ a random sequence of subsets of $[L]$, and $(\ell_t)_t$ a random sequence of functions mapping $\mathcal{V}$ to $\mathbb{R}$ such that $\ell_t(v) \in [0, R_j]$ for all $j \in [L]$, $v \in V_j$, and $t \geq 1$. Accordingly, define*

$$p_1 \coloneqq \operatorname*{arg\,min}_{p \in \Delta_K} \psi_1(p) \quad \text{and for } t \geq 2, \;\; p_t \coloneqq \operatorname*{arg\,min}_{p \in \Delta_K} \left\langle \textstyle\sum_{s \leq t-1} z_s, p \right\rangle + \psi_t(p) \,,$$

where for $t \geq 1$ and $a \in \mathcal{A}$, $z_t(a) \coloneqq \sum_{v \succ a} [\![v \succ A_t, \mathrm{lvl}(v) \in M_t]\!] \frac{\ell_t(v)}{\xi_{\mathrm{lvl}(v)} p_t[v]}$; and for $x \in \mathbb{R}^d$,

$$\psi_t(x) \coloneqq \frac{2\sqrt{c}}{\sqrt{(c-1)}} \sum_{j \in [L]} \frac{R_j}{\sqrt{\xi_j}} \max\left\{ \sqrt{t}, \sqrt{c(c-1)/\xi_j} \right\} \sum_{v \in V_j} x[v] - \sqrt{x[v]}$$

with some constant $c \geq 2$. Assume there exists a filtration $(\mathcal{F}_t)_{t \geq 0}$ such that

(a) $(p_t)_{t \geq 1}$ is predictable w.r.t. $(\mathcal{F}_t)_{t \geq 0}$,
(b) $\mathbb{P}_t(A_t = a) = p_t(a) \;\; \forall a \in \mathcal{A}$,
(c) $\mathbb{P}_t(j \in M_t) = \xi_j \;\; \forall j \in [L]$,
(d) $[\![v \succ A_t]\!] \perp\!\!\!\perp [\![\mathrm{lvl}(v) \in M_t]\!] \mid \mathcal{F}_{t-1} \;\; \forall v \in \mathcal{V}$,
(e) $\sum_{v \succ a} \mathbb{E}_t[\ell_t(v) \mid v \succ A_t, \mathrm{lvl}(v) \in M_t] - \sum_{w \succ a'} \mathbb{E}_t[\ell_t(w) \mid w \succ A_t, \mathrm{lvl}(w) \in M_t] = \mathbb{E}_t[y_t(a) - y_t(a')] \;\; \forall a, a' \in \mathcal{A}$,

where $\mathbb{P}_t(\cdot) = \mathbb{P}(\cdot \mid \mathcal{F}_{t-1})$ and $\mathbb{E}_t[\cdot] = \mathbb{E}[\cdot \mid \mathcal{F}_{t-1}]$. Then, it holds for any horizon $T \geq 1$ that

$$\mathbb{E}\left[\sum_{t=1}^{T} y_t(A_t) \right] - \min_{a \in \mathcal{A}} \mathbb{E}\left[\sum_{t=1}^{T} y_t(a) \right] \leq \frac{4\sqrt{c}}{\sqrt{(c-1)}} \sum_{t=1}^{T} \frac{1}{\sqrt{t}} \sum_{j \in [L]} \frac{R_j}{\sqrt{\xi_j}} \mathbb{E} \sum_{v \in V_j} \left(\sqrt{p_t[v]} - p_t[v] \right) + 3c \sum_{j \in [L]} \frac{R_j}{\xi_j} \sqrt{|V_j|} \,.$$

Proof Let $a^* \in \arg\min_{a\in\mathcal{A}} \mathbb{E}\left[\sum_{t=1}^T y_t(a)\right]$ and let $q \in \Delta_k$ be entirely concentrated on a^*. For brevity, we will use $\widetilde{\ell}_t(v)$ to denote $\mathbb{E}_t[\ell_t(v) \mid v \succ A_t, \mathrm{lvl}(v) \in M_t]$ for $v \in \mathcal{V}$. With this definition, it holds that

$$\begin{aligned}
&\mathbb{E}_t\big[[\![v \succ A_t, \mathrm{lvl}(v) \in M_t]\!]\ell_t(v)\big] \\
&\quad= \mathbb{P}_t(v \succ A_t, \mathrm{lvl}(v) \in M_t)\mathbb{E}_t\big[\ell_t(v)[\![v \succ A_t, \mathrm{lvl}(v) \in M_t]\!] \mid v \succ A_t, \mathrm{lvl}(v) \in M_t\big] \\
&\qquad+ \mathbb{P}_t(v \not\succ A_t \vee \mathrm{lvl}(v) \notin M_t)\mathbb{E}_t\big[\ell_t(v)[\![v \succ A_t, \mathrm{lvl}(v) \in M_t]\!] \mid v \not\succ A_t \vee \mathrm{lvl}(v) \notin M_t\big] \\
&\quad= \mathbb{E}_t\big[[\![v \succ A_t, \mathrm{lvl}(v) \in M_t]\!]\big]\widetilde{\ell}_t(v) = \mathbb{E}_t\big[[\![v \succ A_t]\!]\big]\mathbb{E}_t\big[[\![\mathrm{lvl}(v) \in M_t]\!]\big]\widetilde{\ell}_t(v)\,,
\end{aligned}$$

where the last step holds via the assumed independence of $[\![v \succ A_t]\!]$ and $[\![\mathrm{lvl}(v) \in M_t]\!]$ conditioned on $\mathcal{F}_{t-1}$. Note then that

$$\begin{aligned}
\mathbb{E}\langle z_t, p_t\rangle &= \mathbb{E}\sum_{a\in\mathcal{A}} p_t(a)\sum_{v\succ a}[\![v \succ A_t, \mathrm{lvl}(v) \in M_t]\!]\frac{\ell_t(v)}{\xi_{\mathrm{lvl}(v)}p_t[v]} \\
&= \mathbb{E}\sum_{a\in\mathcal{A}} p_t(a)\sum_{v\succ a}\mathbb{E}_t\big[[\![v \succ A_t, \mathrm{lvl}(v) \in M_t]\!]\ell_t(v)\big]\frac{1}{\xi_{\mathrm{lvl}(v)}p_t[v]} \\
&= \mathbb{E}\sum_{a\in\mathcal{A}} p_t(a)\sum_{v\succ a}\mathbb{E}_t\left[[\![v \succ A_t]\!]\right]\mathbb{E}_t\left[[\![\mathrm{lvl}(v) \in M_t]\!]\right]\frac{\widetilde{\ell}_t(v)}{\xi_{\mathrm{lvl}(v)}p_t[v]} \\
&= \mathbb{E}\sum_{a\in\mathcal{A}} p_t(a)\sum_{v\succ a}\widetilde{\ell}_t(v) \\
&= \mathbb{E}\sum_{v\in\mathcal{V}}\widetilde{\ell}_t(v)\sum_{a\prec v} p_t(a) \\
&= \mathbb{E}\sum_{v\in\mathcal{V}}\widetilde{\ell}_t(v)\mathbb{E}_t[[\![v \succ A_t]\!]] \\
&= \mathbb{E}\sum_{v\in\mathcal{V}}\widetilde{\ell}_t(v)[\![v \succ A_t]\!] = \mathbb{E}\sum_{v\succ A_t}\widetilde{\ell}_t(v)\,,
\end{aligned}$$

where the first equality uses the definition of z_t; the second uses the tower rule and the fact that p_t is deterministic given $\mathcal{F}_{t-1}$; the fourth uses the definition of $p_t[v]$, the assumption that the law of A_t conditioned on $\mathcal{F}_{t-1}$ is p_t, and the assumption that $\mathbb{P}_t(\{\mathrm{lvl}(v) \in M_t\}) = \xi_{\mathrm{lvl}(v)}$; the sixth employs again the fact that the law of A_t conditioned on $\mathcal{F}_{t-1}$ is p_t; and the seventh again uses the tower rule and the fact that $\widetilde{\ell}_t(v)$ is measurable w.r.t. $\mathcal{F}_{t-1}$. In a similar fashion, it holds that

$$\begin{aligned}
\mathbb{E}\langle z_t, q\rangle &= \mathbb{E}\sum_{a\in\mathcal{A}} q(a)\sum_{v\succ a}[\![v \succ A_t, \mathrm{lvl}(v) \in M_t]\!]\frac{\ell_t(v)}{\xi_{\mathrm{lvl}(v)}p_t[v]} \\
&= \mathbb{E}\sum_{a\in\mathcal{A}} q(a)\sum_{v\succ a}\mathbb{E}_t\big[[\![v \succ A_t, \mathrm{lvl}(v) \in M_t]\!]\ell_t(v)\big]\frac{1}{\xi_{\mathrm{lvl}(v)}p_t[v]} \\
&= \mathbb{E}\sum_{a\in\mathcal{A}} q(a)\sum_{v\succ a}\mathbb{E}_t\left[[\![v \succ A_t]\!]\right]\mathbb{E}_t\left[[\![\mathrm{lvl}(v) \in M_t]\!]\right]\frac{\widetilde{\ell}_t(v)}{\xi_{\mathrm{lvl}(v)}p_t[v]} \\
&= \mathbb{E}\sum_{a\in\mathcal{A}} q(a)\sum_{v\succ a}\widetilde{\ell}_t(v)\,.
\end{aligned}$$

Since, by assumption, $\sum_{v \succ a} \widetilde{\ell}_t(v) - \sum_{w \succ a'} \widetilde{\ell}_t(w) = \mathbb{E}_t[y_t(a) - y_t(a')]$ for all $a, a' \in \mathcal{A}$, combining the last two result with an application of the tower rule gets us that

$$\mathbb{E}\langle z_t, p_t - q\rangle = \mathbb{E}\big[y_t(A_t) - \langle y_t, q\rangle\big]\,.$$

Overall, we have shown that

$$\mathbb{E}\left[\sum_{t=1}^{T} y_t(A_t)\right] - \min_{a\in\mathcal{A}} \mathbb{E}\left[\sum_{t=1}^{T} y_t(a)\right] = \mathbb{E}\sum_{t=1}^{T}\langle z_t, p_t - q\rangle\,.$$

We can now invoke Lemma 16 with some $c \geq 2$, $\eta_{t,j} = \frac{\sqrt{\xi_j}}{R_j}\min\left\{\frac{\sqrt{c-1}}{\sqrt{ct}}, \frac{\sqrt{\xi_j}}{c}\right\}$, and $c_{t,j} = [\![j \in M_t]\!]\frac{\ell_t(\mathrm{anc}_j(A_t))}{\xi_j}$ to get that

$$\sum_{t=1}^{T}\langle z_t, p_t - q\rangle \leq 2c\sum_{j\in[L]}\frac{R_j}{\xi_j}\sum_{v\in V_j}\big(\sqrt{p_1[v]} - p_1[v]\big) + \frac{2\sqrt{c}}{\sqrt{(c-1)}}\sum_{t=2}^{T+1}\frac{1}{\sqrt{t}}\sum_{j\in[L]}\frac{R_j}{\sqrt{\xi_j}}\sum_{v\in V_j}\big(\sqrt{p_t[v]} - p_t[v]\big)$$
$$+\frac{\sqrt{c}}{\sqrt{c-1}}\sum_{t=1}^{T}\frac{1}{\sqrt{t}}\sum_{j\in[L]}\frac{\sqrt{\xi_j}}{R_j}\sum_{v\in V_j}\big(p_t[v]\big)^{3/2}\left([\![v = \mathrm{anc}_j(A_t), j \in M_t]\!]\frac{\ell_t(v)}{\xi_j p_t[v]} - [\![j \in M_t]\!]\frac{\ell_t(\mathrm{anc}_j(A_t))}{\xi_j}\right)^2,$$

where we have used that $\widetilde{\eta}_{1,j} = \eta_{1,j}^{-1} = \frac{cR_j}{\xi_j}$, and for $t \geq 2$,

$$\begin{aligned}
\widetilde{\eta}_{t,j} &= \frac{R_j}{\sqrt{\xi_j}}\max\left\{\frac{\sqrt{ct}}{\sqrt{c-1}}, \frac{c}{\sqrt{\xi_j}}\right\} - \frac{R_j}{\sqrt{\xi_j}}\max\left\{\frac{\sqrt{c(t-1)}}{\sqrt{c-1}}, \frac{c}{\sqrt{\xi_j}}\right\} \\
&\leq \frac{\sqrt{c}R_j}{\sqrt{(c-1)\xi_j}}(\sqrt{t} - \sqrt{t-1}) \\
&\leq \frac{\sqrt{c}R_j}{\sqrt{(c-1)\xi_j t}}\,.
\end{aligned}$$

Note also that for every $t \geq 1$, $j \in [L]$, and $v \in V_j$,

$$\begin{aligned}
&[\![v = \mathrm{anc}_j(A_t), j \in M_t]\!]\frac{\ell_t(v)}{\xi_j p_t[v]} - [\![j \in M_t]\!]\frac{\ell_t(\mathrm{anc}_j(A_t))}{\xi_j} \\
&\quad= \frac{[\![j \in M_t]\!]}{\xi_j}\left([\![v = \mathrm{anc}_j(A_t)]\!]\frac{\ell_t(v)}{p_t[v]} - \ell_t(\mathrm{anc}_j(A_t))\right) \\
&\quad\geq -\frac{\ell_t(\mathrm{anc}_j(A_t))}{\xi_j} \\
&\quad\geq -\frac{R_j}{\xi_j} \geq -\frac{1}{c\eta_{t,j}} \geq -\frac{1}{c\eta_{t,j}\sqrt{p_t[v]}}\,;
\end{aligned}$$

thus, the conditions of Lemma 16 are indeed satisfied.

Moving on, it holds that

$$\begin{aligned}
&\mathbb{E}_t \sum_{v\in V_j} \big(p_t[v]\big)^{3/2} \left([\![v = \mathrm{anc}_j(A_t), j \in M_t]\!] \frac{\ell_t(v)}{\xi_j p_t[v]} - [\![j \in M_t]\!] \frac{\ell_t(\mathrm{anc}_j(A_t))}{\xi_j} \right)^2 \\
&\quad = \mathbb{E}_t \frac{[\![j \in M_t]\!]}{\xi_j^2} \sum_{v\in V_j} \big(p_t[v]\big)^{3/2} \left([\![v = \mathrm{anc}_j(A_t)]\!] \frac{\ell_t(v)}{p_t[v]} - \ell_t(\mathrm{anc}_j(A_t)) \right)^2 \\
&\quad = \mathbb{E}_t \frac{[\![j \in M_t]\!] \ell_t^2(\mathrm{anc}_j(A_t))}{\xi_j^2} \left((p_t[\mathrm{anc}_j(A_t)])^{-1/2} - 2(p_t[\mathrm{anc}_j(A_t)])^{1/2} + \textstyle\sum_{v\in V_j} (p_t[v])^{3/2} \right) \\
&\quad \le R_j^2 \mathbb{E}_t \frac{[\![j \in M_t]\!]}{\xi_j^2} \left((p_t[\mathrm{anc}_j(A_t)])^{-1/2} - 2(p_t[\mathrm{anc}_j(A_t)])^{1/2} + \textstyle\sum_{v\in V_j} (p_t[v])^{3/2} \right) \\
&\quad = R_j^2 \mathbb{E}_t \sum_{w\in V_j} \frac{[\![j \in M_t]\!] [\![w \succ A_t]\!]}{\xi_j^2} \left((p_t[w])^{-1/2} - 2(p_t[w])^{1/2} + \textstyle\sum_{v\in V_j} (p_t[v])^{3/2} \right) \\
&\quad = R_j^2 \sum_{w\in V_j} \frac{\mathbb{E}_t[[\![j \in M_t]\!]] \mathbb{E}_t[[\![w \succ A_t]\!]]}{\xi_j^2} \left((p_t[w])^{-1/2} - 2(p_t[w])^{1/2} + \textstyle\sum_{v\in V_j} (p_t[v])^{3/2} \right) \\
&\quad = \frac{R_j^2}{\xi_j} \sum_{w\in V_j} p_t[w] \left((p_t[w])^{-1/2} - 2(p_t[w])^{1/2} + \textstyle\sum_{v\in V_j} (p_t[v])^{3/2} \right) \\
&\quad \le 2\frac{R_j^2}{\xi_j} \sum_{v\in V_j} \sqrt{p_t[v]} - p_t[v] \,,
\end{aligned}$$

where the second equality is an application of the first part of Lemma 15, the fourth equality uses the conditional independence of $[\![w \succ A_t]\!]$ and $[\![\mathrm{lvl}(w) \in M_t]\!]$ given $\mathcal{F}_{t-1}$, and the last inequality is an application of the second part of Lemma 15. Putting everything together, obtain that

$$\begin{aligned}
&\mathbb{E}\left[\sum_{t=1}^T y_t(A_t)\right] - \min_{a\in\mathcal{A}} \mathbb{E}\left[\sum_{t=1}^T y_t(a)\right] \\
&\quad \le \frac{4\sqrt{c}}{\sqrt{(c-1)}} \sum_{t=1}^T \frac{1}{\sqrt{t}} \sum_{j\in[L]} \frac{R_j}{\sqrt{\xi_j}} \mathbb{E} \sum_{v\in V_j} \left(\sqrt{p_t[v]} - p_t[v] \right) + 2c \sum_{j\in[L]} \frac{R_j}{\xi_j} \mathbb{E} \sum_{v\in V_j} \left(\sqrt{p_1[v]} - p_1[v] \right) \\
&\qquad\qquad + \frac{2\sqrt{c}}{\sqrt{(c-1)}} \cdot \frac{1}{\sqrt{T+1}} \sum_{j\in[L]} \frac{R_j}{\sqrt{\xi_j}} \mathbb{E} \sum_{v\in V_j} \left(\sqrt{p_{T+1}[v]} - p_{T+1}[v] \right) .
\end{aligned}$$

The theorem then follows since the sum of the last two terms is no larger than $3c \sum_{j\in[L]} \frac{R_j}{\xi_j} \sqrt{|V_j|}$. ■

Appendix G. Foundational Results

Lemma 18 *Fix a vector $z \in \mathbb{R}^d$ and let $\psi \colon \mathbb{R}^d \to (-\infty, +\infty]$ be a proper and closed convex function. Assume further that* $\mathrm{relint}(\Delta_d) \cap \mathrm{relint}(\mathrm{dom}(\psi)) \neq \emptyset$. *Then, it holds that*

$$\min_{p\in\Delta_d} z^\top p + \psi(p) = \max_{u\in\mathbb{R}^d} -\psi^*(u - z) + \min_i u(i) \,,$$

and that

$$u^* - z \in \partial\psi(p^*)\,,$$

for $p^* \in \arg\min_{p\in\Delta_d} z^\top p + \psi(p)$ *and* $u^* \in \arg\max_{u\in\mathbb{R}^d} -\psi^*(u-z) + \min_i u(i)$*. If, additionally,* ψ *is essentially smooth, then* u^* *is unique,* $p^* \in \Delta_d \cap \operatorname{int}(\operatorname{dom}(\psi))$*, and*

$$u^* - z = \nabla\psi(p^*)\,.$$

If ψ *is also strictly convex on* $\operatorname{int}(\operatorname{dom}(\psi))$*, then* p^* *too is unique.*

Proof Define $\Phi\colon \mathbb{R}^d \to (-\infty, +\infty]$ as

$$\Phi(x) \coloneqq z^\top x + \psi(x)$$

for all $x \in \mathbb{R}^d$. This is a proper and closed convex function satisfying that $\operatorname{dom}(\Phi) = \operatorname{dom}(\psi)$ and $\Phi^*(y) = \psi^*(y - z)$ for any $y \in \mathbb{R}^d$. Define also $g\colon \mathbb{R}^d \to [-\infty, +\infty)$ as

$$g(x) \coloneqq \begin{cases} 0 & \text{if } x \in \Delta_d \\ -\infty & \text{otherwise}\,, \end{cases}$$

which is a closed and proper *concave* function satisfying that $\operatorname{dom}(g) = \Delta_d$, $g^*(y) = \min_i y(i)$ for any $y \in \mathbb{R}^d$, and $\operatorname{dom}(g^*) = \mathbb{R}^d$. Note then that $\operatorname{relint}(\operatorname{dom}(g)) \cap \operatorname{relint}(\operatorname{dom}(\Phi)) \neq \emptyset$ since $\operatorname{relint}(\Delta_d) \cap \operatorname{relint}(\operatorname{dom}(\psi)) \neq \emptyset$ by assumption. Moreover, $\operatorname{relint}(\operatorname{dom}(g^*)) \cap \operatorname{relint}(\operatorname{dom}(\Phi^*)) \neq \emptyset$ since $\operatorname{dom}(g^*) = \mathbb{R}^d$ and $\operatorname{dom}(\Phi^*)$ is not empty (because Φ^* is proper). Hence, Theorem 31.1 in (Rockafellar, 1970) gives that

$$\inf_{x\in\mathbb{R}^d} \Phi(x) - g(x) = \sup_{y\in\mathbb{R}^d} g^*(y) - \Phi^*(y)$$

with both the infimum and the supremum being attained, which entails the first statement of the lemma. Now, fixing $p^* \in \arg\min_{p\in\Delta_d} \Phi(p)$ and $u^* \in \arg\max_{u\in\mathbb{R}^d} -\psi^*(u - z) + \min_i u(i)$, it holds that

$$\begin{aligned}
\Phi(p^*) &= -\psi^*(u^* - z) + \min_i u^*(i) \\
&= \inf_{x\in\mathbb{R}^d} (z - u^*)^\top x + \psi(x) + \min_i u^*(i) \\
&= \min_i u^*(i) + \inf_{x\in\mathbb{R}^d} \Phi(x) - x^\top u^* \\
&\leq \min_i u^*(i) + \Phi(p^*) - (p^*)^\top u^* \leq \Phi(p^*)\,,
\end{aligned}$$

where the last inequality just uses that $p^* \in \Delta_d$. We can then conclude that $p^* \in \arg\min_{x\in\mathbb{R}^d} \psi(x) + x^\top(z - u^*)$, or equivalently

$$p^* \in \arg\max_{x\in\mathbb{R}^d} x^\top(u^* - z) - \psi(x)\,,$$

implying via Theorem 23.5 in (Rockafellar, 1970) that

$$u^* - z \in \partial\psi(p^*)\,, \tag{21}$$

which gives the second statement.

Moving on, if ψ is essentially smooth, Theorem 26.1 in (Rockafellar, 1970) gives that $\partial\psi(x) = \{\nabla\psi(x)\}$ for $x \in \operatorname{int}(\operatorname{dom}(\psi))$ and $\partial\psi(x) = \emptyset$ otherwise. It is then implied by (21) that u^* is unique (since $\partial\psi(p^*)$ is non-empty and must be a singleton), that $p^* \in \operatorname{int}(\operatorname{dom}(\psi))$, and that $u^* - z = \nabla\psi(p^*)$. If ψ is additionally strictly convex on $\operatorname{int}(\operatorname{dom}(\psi))$, Corollary 26.3.1 in (Rockafellar, 1970) gives that $\partial\psi$ is a one-to-one mapping, hence the uniqueness of u^* implies via (21) the uniqueness of p^*. ■

Lemma 19 *For every i in a finite index set $\mathcal{I}$, let $\psi_i \colon \mathbb{R}^d \to (-\infty, +\infty]$ be a proper and closed convex function, and assume that $\cap_{i\in\mathcal{I}} \operatorname{relint}(\operatorname{dom}(\psi_i)) \neq \emptyset$. If every ψ_i is essentially smooth, then so is $\sum_{i\in\mathcal{I}} \psi_i$.*

Proof Let $\psi \coloneqq \sum_{i\in\mathcal{I}} \psi_i$, which is also a proper and closed convex function. Theorem 26.1 in (Rockafellar, 1970) states that a proper and closed convex function $f \colon \mathbb{R}^d \to (\infty, +\infty]$ is essentially smooth if and only if $\partial f(x)$ contains at most one element for any $x \in \mathbb{R}^d$, which hence holds for every ψ_i. Since $\cap_{i\in\mathcal{I}} \operatorname{relint}(\operatorname{dom}(\psi_i)) \neq \emptyset$, (Bertsekas, 2009, Proposition 5.4.6) gives that $\partial\psi(x) = \sum_{i\in\mathcal{I}} \partial\psi_i(x)$ for all $x \in \mathbb{R}^d$. Thus, $\partial\psi(x)$ too contains at most one element, implying the essential smoothness of $\psi(x)$. ■

Lemma 20 *Fix a vector $z \in \mathbb{R}^d$ and a finite index set $\mathcal{I}$. For every $i \in \mathcal{I}$, let $\psi_i \colon \mathbb{R}^d \to (-\infty, +\infty]$ be a proper and closed convex function. Assume further that $\operatorname{relint}(\Delta_d) \cap \big(\cap_{i\in\mathcal{I}} \operatorname{relint}(\operatorname{dom}(\psi_i))\big) \neq \emptyset$. Then, it holds that*

$$\min_{p\in\Delta_d} z^\top p + \sum_{i\in\mathcal{I}} \psi_i(p) = \max_{u\in\mathbb{R}^d} \sup_{\sum_{i\in\mathcal{I}} y_i = u - z} - \sum_{i\in\mathcal{I}} \psi_i^*(y_i) + \min_i u(i)\,,$$

and for any $p^* \in \Delta_d$ minimizing the L.H.S. and $u^* \in \mathbb{R}^d$ maximizing the R.H.S.,

$$\min_{p\in\Delta_d} z^\top p + \sum_{i\in\mathcal{I}} \psi_i(p) = - \sum_{i\in\mathcal{I}} \psi_i^*(g_i) + \min_i u^*(i)\,,$$

for some $(g_i)_{i\in\mathcal{I}}$ satisfying $g_i \in \partial\psi_i(p^*)$ and $\sum_{i\in\mathcal{I}} g_i = u^* - z$. If, additionally, every ψ_i is essentially smooth, then u^* is unique, $p^* \in \Delta_d \cap \operatorname{int}(\cap_{i\in\mathcal{I}} \operatorname{dom}(\psi_i))$, $\sum_{i\in\mathcal{I}} \nabla\psi_i(p^*) = u^* - z$, and

$$\min_{p\in\Delta_d} z^\top p + \sum_{i\in\mathcal{I}} \psi_i(p) = - \sum_{i\in\mathcal{I}} \psi_i^*(\nabla\psi_i(p^*)) + \min_i u^*(i)\,.$$

If, moreover, $\sum_{i\in\mathcal{I}} \psi_i$ is strictly convex on $\operatorname{int}(\cap_{i\in\mathcal{I}} \operatorname{dom}(\psi_i))$, then p^* too is unique.

Proof For brevity, define

$$\psi \coloneqq \sum_{i\in\mathcal{I}} \psi_i\,.$$

This is a convex function that is closed (Bertsekas, 2009, Proposition 1.1.5) and satisfies $\operatorname{dom}(\psi) = \cap_{i\in\mathcal{I}} \operatorname{dom}(\psi_i)$. Hence, our assumption that $\operatorname{relint}(\Delta_d) \cap \big(\cap_{i\in\mathcal{I}} \operatorname{relint}(\operatorname{dom}(\psi_i))\big) \neq \emptyset$ implies that ψ is proper seeing as $\operatorname{dom}(\psi) \neq \emptyset$. Moreover, Theorem 6.5 in (Rockafellar, 1970) gives that

$\operatorname{relint}(\cap_{i\in\mathcal{I}}\operatorname{dom}(\psi_i)) = \cap_{i\in\mathcal{I}}\operatorname{relint}(\operatorname{dom}(\psi_i))$ seeing as $\cap_{i\in\mathcal{I}}\operatorname{relint}(\operatorname{dom}(\psi_i)) \neq \emptyset$. Hence, it holds that $\operatorname{relint}(\Delta_d)\cap\operatorname{relint}(\operatorname{dom}(\psi)) \neq \emptyset$. Lemma 18 then gives that

$$\min_{p\in\Delta_d} z^\top p + \psi(p) = \max_{u\in\mathbb{R}^d} -\psi^*(u-z) + \min_i u(i)\,. \tag{22}$$

Since $\cap_{i\in\mathcal{I}}\operatorname{relint}(\operatorname{dom}(\psi_i)) \neq \emptyset$, Theorem 16.4 in (Rockafellar, 1970) gives that

$$\psi^*(u-z) = \inf_{\sum_{i\in\mathcal{I}} y_i = u-z} \sum_{i\in\mathcal{I}} \psi_i^*(y_i)\,,$$

implying, as sought, that

$$\min_{p\in\Delta_d} z^\top p + \sum_{i\in\mathcal{I}}\psi_i(p) = \max_{u\in\mathbb{R}^d} \sup_{\sum_{i\in\mathcal{I}} y_i = u-z} -\sum_{i\in\mathcal{I}}\psi_i^*(y_i) + \min_i u(i)\,.$$

Fixing $p^* \in \arg\min_{p\in\Delta_d} z^\top p + \psi(p)$ and $u^* \in \arg\max_{u\in\mathbb{R}^d} -\psi^*(u-z) + \min_i u(i)$, Lemma 18 also implies that

$$u^* - z \in \partial\psi(p^*)\,.$$

Hence, by (Bertsekas, 2009, Proposition 5.4.6) and the assumption that $\cap_{i\in\mathcal{I}}\operatorname{relint}(\operatorname{dom}(\psi_i)) \neq \emptyset$, there exists a collection of vectors $(g_i)_{i\in\mathcal{I}}$ each satisfying $g_i \in \partial\psi_i(p^*)$ such that $\sum_{i\in\mathcal{I}} g_i = u^* - z$. We then get that

$$\begin{aligned}
\psi^*(u^*-z) &= (u^*-z)^\top p^* - \psi(p^*) \\
&= \big(\textstyle\sum_{i\in\mathcal{I}} g_i\big)^\top p^* - \psi(p^*) \\
&= \sum_{i\in\mathcal{I}} \big(g_i^\top p^* - \psi_i(p^*)\big) \\
&= \sum_{i\in\mathcal{I}} \psi_i^*(g_i)\,,
\end{aligned}$$

where the first and last equalities follow from (Rockafellar, 1970, Theorem 23.5). Combined with (22), this entails that

$$\min_{p\in\Delta_d} z^\top p + \psi(p) = -\sum_{i\in\mathcal{I}}\psi_i^*(g_i) + \min_i u^*(i)\,. \tag{23}$$

Next, Lemma 19 gives that ψ is essentially smooth when every ψ_i is. If this condition holds, Lemma 18 would additionally give that u^* is unique and that $p^* \in \Delta_d \cap \operatorname{int}(\cap_{i\in\mathcal{I}}\operatorname{dom}(\psi_i))$. The latter implies that $p^* \in \operatorname{int}(\operatorname{dom}(\psi_i))$ for every $i \in \mathcal{I}$, which entails via Theorem 26.1 in (Rockafellar, 1970) that $\partial\psi_i(p^*) = \{\nabla\psi_i(p^*)\}$. Hence, g_i in (23) becomes $\nabla\psi_i(p^*)$ and $\sum_{i\in\mathcal{I}}\nabla\psi_i(p^*) = u^* - z$. Finally, Lemma 18 implies that p^* is unique if it also holds that $\sum_i \psi_i$ is strictly convex in $\operatorname{int}(\cap_{i\in\mathcal{I}}\operatorname{dom}(\psi_i))$. ■

For a convex function $f\colon\mathbb{R}\to(-\infty,+\infty]$ and a vector $h\in\mathbb{R}^d$, we denote by $f\circ h^\top\colon\mathbb{R}^d\to(-\infty,+\infty]$ the function given by

$$(f\circ h^\top)(x) \coloneqq f(h^\top x)\,.$$

Whereas the function $h\circ f\colon\mathbb{R}^d\to(-\infty,+\infty]$ is given by

$$(h\circ f)(x) \coloneqq \inf_{x=\alpha h} f(\alpha)\,.$$

Lemma 21 *For a proper convex function $f\colon \mathbb{R} \to (-\infty, +\infty]$ and a non-zero vector $h \in \mathbb{R}^d$, it holds that*

$$(f \circ h^\top)^* = h \circ f^* \,.$$

Proof Note that any vector $y \in \mathbb{R}^d$ satisfies $y = \alpha h + y'$ for some $\alpha \in \mathbb{R}$ and $y' \in \mathbb{R}^d$ such that $h^\top y' = 0$. Then, starting from the definition of the conjugate, we have that

$$\begin{aligned}
(f \circ h^\top)^*(y) &= \sup_{x \in \mathbb{R}^d} x^\top y - f(h^\top x) \\
&= \sup_{x \in \mathbb{R}^d} x^\top (\alpha h + y') - f(h^\top x) \\
&= \sup_{\beta \in \mathbb{R},\, x' \in \mathbb{R}^d \colon\, h^\top x' = 0} (\beta h + x')^\top (\alpha h + y') - f(h^\top (\beta h + x')) \\
&= \sup_{\beta \in \mathbb{R},\, x' \in \mathbb{R}^d \colon\, h^\top x' = 0} \alpha\beta \|h\|^2 + x'^\top y' - f(\beta \|h\|^2) \\
&= f^*(\alpha) + \sup_{x' \in \mathbb{R}^d \colon\, h^\top x' = 0} x'^\top y' \\
&= \inf_{y = \alpha h} f^*(\alpha) = (h \circ f^*)(y) \,,
\end{aligned}$$

where the penultimate equality uses that the supremum is 0 if y and h are collinear and $+\infty$ otherwise. In the latter case $\{\alpha \in \mathbb{R} \colon y = \alpha h\}$ is empty and $(h \circ f^*)(y)$ is vacuously $+\infty$, while in the former, said set is a singleton. ■

Lemma 22 *For a proper convex function $f\colon \mathbb{R} \to (-\infty, +\infty]$ and a non-zero vector $h \in \mathbb{R}^d$, it holds for every $x \in \mathbb{R}^d$ that*

$$\partial (f \circ h^\top)(x) = h \partial f(h^\top x) \,.$$

Proof This follows directly from Theorem 23.9 in (Rockafellar, 1970) since the range of $h^\top$ is $\mathbb{R}$, and so, it automatically has a non-empty intersection with $\operatorname{relint}(\operatorname{dom}(f))$ (Rockafellar, 1970, Theorem 6.2). ■

Lemma 23 *Let $f\colon \mathbb{R} \to (-\infty, +\infty]$ be a proper and closed convex function, and let $h \in \mathbb{R}^d$ be a non-zero vector. If f is essentially smooth, then so is $f \circ h^\top$.*

Proof Firstly, $f \circ h^\top$ is also a closed convex function via Proposition 1.1.4 in (Bertsekas, 2009). It is also proper since the range of $h^\top$ is the entirety of $\mathbb{R}$. Theorem 26.1 in (Rockafellar, 1970) states that a proper and closed convex function is essentially smooth if and only if the sub-differential at any point contains at most one element. Since f is essentially smooth and, by Lemma 22, every $x \in \mathbb{R}^d$ satisfies

$$\partial (f \circ h^\top)(x) = h \partial f(h^\top x) \,,$$

it follows that $(f \circ h^\top)$ too is essentially smooth. ■

Lemma 24 *For every i in a finite index set $\mathcal{I}$, let $f_i\colon \mathbb{R} \to (-\infty, +\infty]$ be a proper convex function and $h_i \in \mathbb{R}^d$ be a non-zero vector. Assume that $\operatorname{int}(\cap_{i\in\mathcal{I}} \operatorname{dom}(f_i \circ h_i^\top)) \neq \emptyset$ and that $\operatorname{span}\{h_i\}_{i\in\mathcal{I}} = \mathbb{R}^d$. Then, if every f_i is strictly convex in $\operatorname{int}(\operatorname{dom}(f_i))$, it holds that $\sum_{i\in\mathcal{I}} f_i \circ h_i^\top$ is strictly convex in $\operatorname{int}(\cap_{i\in\mathcal{I}} \operatorname{dom}(f_i \circ h_i^\top))$.*

Proof Note that $\operatorname{dom}(f_i \circ h_i^\top) = \{x \in \mathbb{R}^d\colon h_i^\top x \in \operatorname{dom}(f_i)\}$. Define $\psi \coloneqq \sum_{i\in\mathcal{I}} f_i \circ h_i^\top$. Then, $\operatorname{dom}(\psi) = \cap_{i\in\mathcal{I}} \operatorname{dom}(f_i \circ h_i^\top)$ and ψ is a proper convex function. Note, moreover, that $x \mapsto h_i^\top x$ maps vectors in $\operatorname{int}(\operatorname{dom}(f_i \circ h_i^\top))$ to $\operatorname{int}(\operatorname{dom}(f_i))$. Now, for any $\alpha \in (0,1)$ and $x, y \in \operatorname{int}(\operatorname{dom}(\psi))$ such that $x \neq y$, we have that

$$\begin{aligned}
\psi(\alpha x + (1-\alpha)y) &= \sum_{i\in\mathcal{I}} f_i(h_i^\top(\alpha x + (1-\alpha)y)) \\
&= \sum_{i\in\mathcal{I}} f_i(\alpha h_i^\top x + (1-\alpha)h_i^\top y) \\
&< \sum_{i\in\mathcal{I}} \alpha f_i(h_i^\top x) + (1-\alpha) f_i(h_i^\top y) \\
&= \alpha\psi(x) + (1-\alpha)\psi(y)\,,
\end{aligned}$$

where the inequality holds since $h_i^\top x, h_i^\top y \in \operatorname{int}(\operatorname{dom}(f_i \circ h_i^\top))$ for all $i \in \mathcal{I}$ and $h_i^\top x \neq h_i^\top y$ for at least one $i \in \mathcal{I}$ since $x - y \neq 0$ and $\operatorname{span}\{h_i\}_{i\in\mathcal{I}} = \mathbb{R}^d$. We conclude then that ψ is strictly convex in $\operatorname{int}(\operatorname{dom}(\psi))$ as required. ■

Lemma 25 *Fix a vector $z \in \mathbb{R}^d$ and a finite index set $\mathcal{I}$. For every $i \in \mathcal{I}$, let $h_i \in \mathbb{R}^d$ be a non-zero vector and $f_i\colon \mathbb{R} \to (-\infty, +\infty]$ a proper and closed convex function. Assume that $\operatorname{relint}(\Delta_d) \cap \left(\cap_{i\in\mathcal{I}} \operatorname{relint}(\{x \in \mathbb{R}^d\colon h_i^\top x \in \operatorname{dom}(f_i)\})\right) \neq \emptyset$. Then, it holds that*

$$\min_{p\in\Delta_d} z^\top p + \sum_{i\in\mathcal{I}} f_i(h_i^\top p) = \max_{u\in\mathbb{R}^d} \sup_{\sum_{i\in\mathcal{I}} \alpha_i h_i = u - z} -\sum_{i\in\mathcal{I}} f_i^*(\alpha_i) + \min_i u(i)\,,$$

and for any $p^* \in \Delta_d$ minimizing the L.H.S. and $u^* \in \mathbb{R}^d$ maximizing the R.H.S.,

$$\min_{p\in\Delta_d} z^\top p + \sum_{i\in\mathcal{I}} f_i(h_i^\top p) = -\sum_{i\in\mathcal{I}} f_i^*(\alpha_i^*) + \min_i u^*(i)\,,$$

for some $(\alpha_i^*)_{i\in\mathcal{I}}$ satisfying $\alpha_i^* \in \partial f_i(h_i^\top p^*)$ and $\sum_{i\in\mathcal{I}} \alpha_i^* h_i = u^* - z$. If, additionally, every f_i is essentially smooth, then u^* is unique, $p^* \in \Delta_d \cap \operatorname{int}(\cap_{i\in\mathcal{I}}\{x \in \mathbb{R}^d\colon h_i^\top x \in \operatorname{dom}(f_i)\})$, $\sum_{i\in\mathcal{I}} f_i'(h_i^\top p^*) h_i = u^* - z$, and

$$\min_{p\in\Delta_d} z^\top p + \sum_{i\in\mathcal{I}} f_i(h_i^\top p) = -\sum_{i\in\mathcal{I}} f_i^*(f_i'(h_i^\top p^*)) + \min_i u^*(i)\,.$$

If, moreover, $\operatorname{span}\{h_i\}_{i\in\mathcal{I}} = \mathbb{R}^d$ and every f_i is strictly convex on $\operatorname{int}(\operatorname{dom}(f_i))$, then p^* too is unique.

Proof Define $\psi_i \coloneqq f_i \circ h_i^\top$. Notice then that $\operatorname{dom}(\psi_i) = \{x \in \mathbb{R}^d \colon h_i^\top x \in \operatorname{dom}(f_i)\}$. Moreover, ψ_i is a closed and proper convex function. Hence, our assumption that $\operatorname{relint}(\Delta_d) \cap \big(\cap_{i\in\mathcal{I}} \operatorname{relint}(\{x \in \mathbb{R}^d \colon h_i^\top x \in \operatorname{dom}(f_i)\})\big) \neq \emptyset$ implies via Lemma 20 that

$$\min_{p\in\Delta_d} z^\top p + \sum_{i\in\mathcal{I}} \psi_i(p) = \max_{u\in\mathbb{R}^d} \sup_{\sum_{i\in\mathcal{I}} y_i = u - z} -\sum_{i\in\mathcal{I}} \psi_i^*(y_i) + \min_i u(i)\,.$$

Using this, the definition of ψ_i, and Lemma 21, we obtain that

$$\begin{aligned}\min_{p\in\Delta_d} z^\top p + \sum_{i\in\mathcal{I}} f_i(h_i^\top p) &= \max_{u\in\mathbb{R}^d} \sup_{\sum_{i\in\mathcal{I}} y_i = u - z} -\sum_{i\in\mathcal{I}} \inf_{y_i = \alpha_i h_i} f_i^*(\alpha_i) + \min_i u(i) \\ &= \max_{u\in\mathbb{R}^d} \sup_{\sum_{i\in\mathcal{I}} \alpha_i h_i = u - z} -\sum_{i\in\mathcal{I}} f_i^*(\alpha_i) + \min_i u(i)\,.\end{aligned}$$

For any

$$\begin{aligned}p^* \in \arg\min_{p\in\Delta_d} z^\top p + \sum_{i\in\mathcal{I}} f_i(h_i^\top p) \\ \text{and} \quad u^* \in \arg\max_{u\in\mathbb{R}^d} \sup_{\sum_{i\in\mathcal{I}} \alpha_i h_i = u - z} -\sum_{i\in\mathcal{I}} f_i^*(\alpha_i) + \min_i u(i)\,,\end{aligned}$$

Lemma 20 also gives that

$$\min_{p\in\Delta_d} z^\top p + \sum_{i\in\mathcal{I}} f_i(h_i^\top p) = -\sum_{i\in\mathcal{I}} \psi_i^*(g_i) + \min_i u^*(i)\,,$$

for some $(g_i)_{i\in\mathcal{I}}$ satisfying $g_i \in \partial\psi_i(p^*)$ and $\sum_{i\in\mathcal{I}} g_i = u^* - z$. On the other hand, the definition of ψ_i and Lemma 22 give that

$$\partial\psi_i(p^*) = \partial(f_i \circ h_i^\top)(p^*) = h_i \partial f_i(h_i^\top p^*)\,,$$

implying the existence of real weights $(\alpha_i^*)_{i\in\mathcal{I}}$ such that $g_i = \alpha_i^* h_i$ and $\alpha_i^* \in \partial f_i(h_i^\top p^*)$. Hence, Lemma 21 gives that

$$\psi_i^*(g_i) = \psi_i^*(\alpha_i^* h_i) = \inf_{\alpha_i^* h_i = \alpha h_i} f_i^*(\alpha) = f_i^*(\alpha_i^*)\,,$$

using in the end that $h_i \neq \mathbf{0}$. We have thus shown that

$$\min_{p\in\Delta_d} z^\top p + \sum_{i\in\mathcal{I}} f_i(h_i^\top p) = -\sum_{i\in\mathcal{I}} f_i^*(\alpha_i^*) + \min_i u^*(i)\,, \tag{24}$$

for some $(\alpha_i^*)_{i\in\mathcal{I}}$ satisfying $\alpha_i^* \in \partial f_i(h_i^\top p^*)$ and $\sum_{i\in\mathcal{I}} \alpha_i^* h_i = u^* - z$.

Next, if f_i is essentially smooth, then so is $f_i \circ h_i^\top$ by Lemma 23. Under that condition, Lemma 20 gives that u^* is unique and that $p^* \in \Delta_d \cap \operatorname{int}(\cap_{i\in\mathcal{I}}\{x \in \mathbb{R}^d \colon h_i^\top x \in \operatorname{dom}(f_i)\})$. The latter implies that $p^* \in \operatorname{int}(\{x \in \mathbb{R}^d \colon h_i^\top x \in \operatorname{dom}(f_i)\})$ for every $i \in \mathcal{I}$, which entails that $h_i^\top p^* \in \operatorname{int}(\operatorname{dom}(f_i))$. The essential smoothness of f_i then entails via Theorem 26.1 in (Rockafellar, 1970) that $\partial f_i(h_i^\top p^*) = \{f_i'(h_i^\top p^*)\}$. Hence, α_i^* in (24) becomes $f_i'(h_i^\top p^*)$ and we get that $\sum_{i\in\mathcal{I}} f_i'(h_i^\top p^*) h_i = u^* - z$. Finally, Lemma 24 gives that $\sum_i \psi_i = \sum_i f_i \circ h_i^\top$ is strictly convex in

the interior of its domain whenever it holds that $\operatorname{span}\{h_i\}_{i\in\mathcal{I}} = \mathbb{R}^d$ and every f_i is strictly convex in the interior of its domain (which is not empty via essential smoothness). Thus, Lemma 20 implies under these conditions that p^* is also unique. ■

For a function $f\colon \mathbb{R}^d \to (-\infty, +\infty]$ that is differentiable in the interior of its domain, the Bregman divergence with respect to f of $x \in \mathbb{R}^d$ from $y \in \operatorname{int}(\operatorname{dom}(f)$ is defined as

$$D_f(x\|y) \coloneqq f(x) - f(y) - \langle \nabla f(y), x - y\rangle \,.$$

Proposition 26 *Fix a finite index set $\mathcal{I}$ and associate to every $i \in \mathcal{I}$ a non-zero vector $h_i \in \mathbb{R}^d$ and a weight $\eta_i > 0$. Let $f\colon \mathbb{R} \to (-\infty, +\infty]$ be a convex function that is proper, closed, and essentially smooth. Assume also that $\operatorname{relint}(\Delta_d) \cap \big(\cap_{i\in\mathcal{I}} \operatorname{relint}(\{x \in \mathbb{R}^d\colon h_i^\top x \in \operatorname{dom}(f)\})\big) \neq \emptyset$. Fix $y, z \in \mathbb{R}^d$ and let*

$$p^* \in \operatorname*{arg\,min}_{p\in\Delta_d} y^\top p + \sum_{i\in\mathcal{I}} \frac{1}{\eta_i} f(h_i^\top p) \,.$$

Then,

$$\begin{aligned}(y+z)^\top p^* + \sum_{i\in\mathcal{I}} \frac{1}{\eta_i} f(h_i^\top p^*) - \min_{p\in\Delta_d}\left\{(y+z)^\top p + \sum_{i\in\mathcal{I}} \frac{1}{\eta_i} f(h_i^\top p)\right\} \\ \leq \inf_{c,\{\alpha_i\}_{i\in\mathcal{I}}\colon \sum_{i\in\mathcal{I}} \alpha_i h_i = z - c\mathbf{1}} \sum_{i\in\mathcal{I}} \frac{1}{\eta_i} D_{f^*}\big(f'(h_i^\top p^*) - \eta_i\alpha_i \big\| f'(h_i^\top p^*)\big) \,.\end{aligned}$$

Proof Fix any $p \in \Delta_d$ and constant $c \in \mathbb{R}$. Since $\langle c\mathbf{1}, p^* - p\rangle = 0$, we get that

$$\begin{aligned}(y+z)^\top p^* + \sum_{i\in\mathcal{I}} \frac{1}{\eta_i} f(h_i^\top p^*) - (y+z)^\top p - \sum_{i\in\mathcal{I}} \frac{1}{\eta_i} f(h_i^\top p) \\ = (y+z)^\top (p^* - p) + \sum_{i\in\mathcal{I}} \frac{1}{\eta_i} f(h_i^\top p^*) - \sum_{i\in\mathcal{I}} \frac{1}{\eta_i} f(h_i^\top p) \\ = (y+z-c\mathbf{1})^\top (p^* - p) + \sum_{i\in\mathcal{I}} \frac{1}{\eta_i} f(h_i^\top p^*) - \sum_{i\in\mathcal{I}} \frac{1}{\eta_i} f(h_i^\top p) \,.\end{aligned}$$

It can be readily seen that $f_i \coloneqq (1/\eta_i) f$ satisfies the same properties of f that are listed in the statement of the proposition. Hence, Lemma 25 gives that

$$\begin{aligned}y^\top p^* + \sum_{i\in\mathcal{I}} \frac{1}{\eta_i} f(h_i^\top p^*) &= -\sum_{i\in\mathcal{I}} f_i^*(f_i'(h_i^\top p^*)) + \min_i u_{p^*}(i) \\ &= -\sum_{i\in\mathcal{I}} \frac{1}{\eta_i} f^*(\eta_i f_i'(h_i^\top p^*)) + \min_i u_{p^*}(i) \\ &= -\sum_{i\in\mathcal{I}} \frac{1}{\eta_i} f^*(f'(h_i^\top p^*)) + \min_i u_{p^*}(i)\end{aligned}$$

for some $u_{p^*} \in \mathbb{R}^d$, which satisfies that $\sum_{i\in\mathcal{I}} (1/\eta_i) f'(h_i^\top p^*) h_i = u_{p^*} - y$. Lemma 25 also implies that $p^* \in \operatorname{int}(\cap_{i\in\mathcal{I}}\{x \in \mathbb{R}^d\colon h_i^\top x \in \operatorname{dom}(f)\})$, thus, as a consequence, $h_i^\top p^* \in \operatorname{int}(\operatorname{dom}(f))$ for

every $i \in \mathcal{I}$. At the same time, another implication of Lemma 25 is that

$$\begin{aligned}
&-(y+z-c\mathbf{1})^\top p - \sum_{i\in\mathcal{I}} \frac{1}{\eta_i} f(h_i^\top p) \\
&\qquad \leq \min_{u\in\mathbb{R}^d} \inf_{\sum_{i\in\mathcal{I}} \alpha_i h_i = u-z-y+c\mathbf{1}} \sum_{i\in\mathcal{I}} f_i^*(\alpha_i) - \min_i u(i) \\
&\qquad \leq -\min_i u_{p^*}(i) + \inf_{\sum_{i\in\mathcal{I}} \alpha_i h_i = u_{p^*}-z-y+c\mathbf{1}} \sum_{i\in\mathcal{I}} f_i^*(\alpha_i) \\
&\qquad = -\min_i u_{p^*}(i) + \inf_{\sum_{i\in\mathcal{I}} \alpha_i h_i = u_{p^*}-z-y+c\mathbf{1}} \sum_{i\in\mathcal{I}} \frac{1}{\eta_i} f^*(\eta_i\alpha_i) \\
&\qquad = -\min_i u_{p^*}(i) + \inf_{\sum_{i\in\mathcal{I}} \alpha_i h_i = -z+c\mathbf{1}} \sum_{i\in\mathcal{I}} \frac{1}{\eta_i} f^*(f'(h_i^\top p^*) + \eta_i\alpha_i) \\
&\qquad = -\min_i u_{p^*}(i) + \inf_{\sum_{i\in\mathcal{I}} \alpha_i h_i = z-c\mathbf{1}} \sum_{i\in\mathcal{I}} \frac{1}{\eta_i} f^*(f'(h_i^\top p^*) - \eta_i\alpha_i)\,,
\end{aligned}$$

where the third equality uses that $\sum_{i\in\mathcal{I}} (1/\eta_i) f'(h_i^\top p^*) h_i = u_{p^*} - y$.

Combining the results derived thus far we get that

$$\begin{aligned}
&(y+z)^\top p^* + \sum_{i\in\mathcal{I}} \frac{1}{\eta_i} f(h_i^\top p^*) - (y+z)^\top p - \sum_{i\in\mathcal{I}} \frac{1}{\eta_i} f(h_i^\top p) \\
&\leq (z-c\mathbf{1})^\top p^* + \inf_{\sum_{i\in\mathcal{I}} \alpha_i h_i = z-c\mathbf{1}} \sum_{i\in\mathcal{I}} \frac{1}{\eta_i} f^*(f'(h_i^\top p^*) - \eta_i\alpha_i) - \frac{1}{\eta_i} f^*(f'(h_i^\top p^*)) \\
&= \inf_{\sum_{i\in\mathcal{I}} \alpha_i h_i = z-c\mathbf{1}} (z-c\mathbf{1})^\top p^* + \sum_{i\in\mathcal{I}} \frac{1}{\eta_i} f^*(f'(h_i^\top p^*) - \eta_i\alpha_i) - \frac{1}{\eta_i} f^*(f'(h_i^\top p^*)) \\
&= \inf_{\sum_{i\in\mathcal{I}} \alpha_i h_i = z-c\mathbf{1}} \sum_{i\in\mathcal{I}} \alpha_i h_i^\top p^* + \sum_{i\in\mathcal{I}} \frac{1}{\eta_i} f^*(f'(h_i^\top p^*) - \eta_i\alpha_i) - \frac{1}{\eta_i} f^*(f'(h_i^\top p^*)) \\
&= \inf_{\sum_{i\in\mathcal{I}} \alpha_i h_i = z-c\mathbf{1}} \sum_{i\in\mathcal{I}} \frac{1}{\eta_i} \Big(\eta_i\alpha_i h_i^\top p^* + f^*(f'(h_i^\top p^*) - \eta_i\alpha_i) - f^*(f'(h_i^\top p^*)) \Big) \\
&= \inf_{\sum_{i\in\mathcal{I}} \alpha_i h_i = z-c\mathbf{1}} \sum_{i\in\mathcal{I}} \frac{1}{\eta_i} \Big(\eta_i\alpha_i f^{*\prime}(f'(h_i^\top p^*)) + f^*(f'(h_i^\top p^*) - \eta_i\alpha_i) - f^*(f'(h_i^\top p^*)) \Big) \\
&= \inf_{\sum_{i\in\mathcal{I}} \alpha_i h_i = z-c\mathbf{1}} \sum_{i\in\mathcal{I}} \frac{1}{\eta_i} D_{f^*}\big(f'(h_i^\top p^*) - \eta_i\alpha_i \big\| f'(h_i^\top p^*) \big)\,,
\end{aligned}$$

where the fourth equality follows from Theorem 26.5 in (Rockafellar, 1970) and the fact that $h_i^\top p^* \in \operatorname{int}(\operatorname{dom}(f))$ for every $i \in \mathcal{I}$. The sought result now follows after taking the infimum over c and the minimum over p. ■