OPT2023: 15th Annual Workshop on Optimization for Machine Learning

Decentralized Learning Dynamics in the Gossip Model

John Lazarsfeld JOHN.LAZARSFELD@YALE.EDU
Yale University

Dan Alistarh DAN.ALISTARH@IST.AC.AT
IST Austria

Abstract

We study a distributed multi-armed bandit setting among a population of n memory-constrained nodes in the gossip model: at each round, every node locally adopts one of m arms, observes a reward drawn from the arm's (adversarially chosen) distribution, and then communicates with a randomly sampled neighbor, exchanging information to determine its policy in the next round. We introduce and analyze several families of dynamics for this task that are *decentralized*: each node's decision is entirely local and depends only on its most recently obtained reward and that of the neighbor it sampled. We show a connection between the global evolution of these decentralized dynamics with a certain class of *"zero-sum" multiplicative weight update* algorithms, and we develop a general framework for analyzing the population-level regret of these natural protocols. Using this framework, we derive sublinear regret bounds under a wide range of parameter regimes (i.e., the size of m and n) in an adversarial reward setting (where the mean of each arm's distribution can vary over time), when the number of rounds T is at most logarithmic in n. Further, we show that these protocols can approximately optimize convex functions over the simplex when the reward distributions are generated from a stochastic gradient oracle.

1. Introduction

Multi-armed bandits are a powerful and general abstraction in online learning [9, 20, 29], and recently there has been significant interest in *distributed, multi-player variants*, in which multiple agents can sample in parallel and can coordinate under limited communication [2, 4, 6, 8, 12, 14, 18, 19, 22, 23, 31, 34]. However, one setting that has received significantly less attention is the *decentralized setting*, e.g. [21], in which individual nodes lack a global view of the set of arms, have limited memory, and can only exchange information via random, direct exchanges.

In decentralized learning, the goal of the system would be to collectively limit regret at the *population level* by ensuring that nodes consistently choose better arms at each round. This model is well-motivated by distributed settings in which it is impractical for a single node to obtain global information about the system. We briefly introduce the setting and objective as follows:

Problem Setting *Consider a population of n nodes distributed over a complete communication graph. The nodes interact with an m-armed bandit instance over a sequence of T rounds, each of which is structured as follows:*

(i) ***Arm Adoption****: At the start of each round t, each node $u \in [n]$ must adopt one of the m arms.*

(ii) ***Reward Generation and Observation****: Then, each arm j generates a single stochastic reward $g_j^t \sim \nu_j^t$, where ν_j^t is a distribution supported on $[-\sigma, \sigma]$ with mean $\mu_j^t \in [-1, 1]$, for some $1 \leq \sigma \leq 10$. Every node adopting arm j subsequently observes the reward g_j^t.*

(iii) ***Communication****: Then, every node u simultaneously samples a neighbor to receive information from, uniformly at random. Each node subsequently uses this interaction to inform its adoption strategy at round $t+1$, following a fixed local (randomized) protocol.*

Let $\mathbf{p}^t := (p_1^t, \dots, p_m^t) \in \Delta_m$ denote the distribution[1] *specifying the fraction of the population adopting each arm j at round t, and define $\mathbf{g}^t := (g_1^t, \dots, g_m^t)$ and $\boldsymbol{\mu}^t := (\mu_1^t, \dots, \mu_m^t)$. Then the objective of the population is to minimize its expected population-level regret $R(T)$,*[2]

$$R(T) \;:=\; \max_{j\in[m]} \sum_{t\in[T]} \mu_j^t \;-\; \sum_{t\in[T]} \mathbb{E}[\langle \mathbf{p}^t, \mathbf{g}^t \rangle] \,. \tag{1}$$

Discussion and Related Models From a node-local perspective, this is a setting of an adversarial m-armed bandit. However, from a global perspective, the full reward vector $\mathbf{g}^t$ is distributed over the population at every round, making it a decentralized version of *prediction with expert advice* [3], which is reflected in the objective. Other than the boundedness conditions, the setting makes no assumptions on how the rewards g_j^t are generated; in particular we assume an **adversarial reward setting**, where each $\boldsymbol{\mu}^t$ can change over rounds. Communication occurs according to the standard synchronous gossip model with uniform neighbor sampling [7, 28]. As such, our model is related to the *decentralizated optimization model*, which has become popular when studying variants of SGD-based optimization [15, 17, 33]. Moreover, processes within this model are closely related to distributed opinion dynamics [5] (which are used to model complex behavior in distributed systems and also social and biological settings) and to evolutionary game theory [1, 27].

The most closely related models were studied by Celis et al. [10], Su et al. [30], and Sankararaman et al. [25]. However, all these works consider only a stationary setting where rewards come from fixed-mean Bernoulli distributions. Additionally, they assume that (a) at any round a node can choose *not* to adopt an arm, and that either (b) a node can alternate between strategies each round, and, e.g., choose an arm uniformly at random with some small probability [10, 30], (c) that node communication is implicitly performed through some centralized coordination to account for nodes that made no adoption choice at the current round [10], or that (d) a node can remember its adoption history from multiple prior rounds, and where each node adopting arm j at round t generates its own, independent reward from the distribution ν_j^t [25].

In contrast, our focus is to design and analyze extremely *simple, memory-constrained dynamics* in which (i) a node can remember only its most recent arm choice and reward observation (ii) every node runs an identical, fixed protocol that applies the same simple decision rule at every round, and that (iii) nodes *cannot* choose to adopt one of m arms uniformly at random (since at the end of each round, each node is only aware of their most recent adoption decision, and that of the neighbor they communicated with). In other words: we desire dynamics that are fully *decentralized*, and that are also robust to adversarially chosen rewards.

Our contributions We introduce and analyze several families of local dynamics satisfying properties (i), (ii), and (iii) above, and we obtain bounds on population regret $R(T)$ that scale sublinearly with T. Informally, for general *adversarial rewards*, our dynamics obtain average regret

$$\frac{1}{T}R(T) \;\le\; O\left(\sqrt{\frac{\log m}{T}}\right) + \widetilde{O}\left(\frac{\sigma m}{n^{\epsilon}} + \frac{\sigma m}{n^{c}}\right) ,$$

1. We write $\Delta_m := \{\mathbf{p} \in \mathbb{R}^m : \|\mathbf{p}\|_1 = 1\}$ to denote the probability simplex over m coordinates.
2. In standard bandit settings, this quantity is often called *pseudo-regret* [9, 20]. We use the term *regret* for clarity and note that this is the same objective considered in the related work of Celis et al. [10].

for any $c \geq 1$, $n \geq 3c \log n$, and $T \leq (\frac{1}{2} - \epsilon) \log n$ rounds, for any $\epsilon \in (0, \frac{1}{2})$.

Here, the $\widetilde{O}(\cdot)$ surpresses (for readability) lower-order logarithmic dependencies on m and n, and by virtue of the analysis framework we introduce, this regret bound (stated formally in Theorem 2.3) is decomposed into an *approximation error* (where we bound the regret of a "smoother" version of the induced sequence $\{\mathbf{p}^t\}$) and an *estimation error* (where we pay for the error introduced by this coupling). In our bound, the approximation error matches the known optimal regret in the (centralized) prediction with expert advice setting [3], while the estimation error can be interpreted as a *cost of decentralization*: in order to achieve sublinear regret (or equivalently, vanishing average regret), the population size n must grow sufficiently large with respect to m, and thus the regret can generally be sharpened with larger populations. Additionally, in this adversarial reward setting, we show under the constraints of the problem setting that regret can grow linearly with T if the number of rounds can be arbitrarily large. For this reason, our bounds constrain T to grow at most logarithmically in the population size n.

Roughly speaking, we obtain these bounds by analyzing the evolution of the sequences $\{\mathbf{p}^t\}$ induced by our families of dynamics. Surprisingly, we show for each dynamics that the adoption mass p_j^t evolves (in conditional expectation) by multiplicative factors of the form $(1 + F_j(\mathbf{p}^t, \mathbf{g}^t))$ for each arm j, and where each F_j is a function depending on $\mathbf{p}^t$ and $\mathbf{g}^t$ that collectively satisfy the key "zero-sum" property of $\sum_{j \in [m]} p_j^t \cdot F_j(\mathbf{p}^t, \mathbf{g}^t) = 0$. We more generally relate processes of this form to a class of (centralized) *zero-sum* multiplicative weights update (MWU) algorithms, and we derive bounds on their regret that may be of independent interest. This connection is then leveraged to establish a general analysis framework for bounding the regret of the original process $\{\mathbf{p}^t\}$.

Finally, using the known connections between (online) convex optimization and the standard, centralized MWU algorithm (and related processes like mirror descent and the exponentiated gradient method) [3, 11, 13], we use our dynamics and analysis framework to obtain expected error rates for optimizing a convex function $f : \Delta_m \to \mathbb{R}$ by assuming the reward sequence $\{\mathbf{g}^t\}$ is generated by a stochastic gradient oracle. For this, we give an error rate at the *average* iterate $\widetilde{\mathbf{p}} := \frac{1}{T} \sum_{t \in [T]} \mathbf{p}^t$ induced by our protocols that matches (up to some constant factors) the regret bound from the adversarial reward setting above. This result is given formally in Theorem 2.4.

2. Technical Overview of Results

Notation and Other Preliminaries Throughout, we deal with multiple sequences of vectors indexed over rounds $t \in [T]$, for which we use the short hand notation $\{\mathbf{p}^t\} := \mathbf{p}^0, \mathbf{p}^1, \ldots, \mathbf{p}^t$. We often compute expectation (resp., probabilities) conditioned on two sequences $\{\mathbf{p}^t\}$ and $\{\mathbf{g}^t\}$ (or $\{\mathbf{q}^t\}$ and $\{\mathbf{g}^t\}$) simultaneously, and we denote this double conditioning by $\mathbb{E}_t[\cdot]$. When we wish to condition just on a single vector $\mathbf{p}^t$, we will usually write $\mathbb{E}_{\mathbf{p}^t}[\cdot]$. Given $\mathbf{p} = (p_1, \ldots, p_m)$, we write $\mathbb{E}[\mathbf{p}]$ to denote the vector $(\mathbb{E}[p_1], \ldots, \mathbb{E}[p_m])$. Throughout, we use the fact that for a random variable x, if $x \leq \alpha$ with probability at least $1 - \gamma$, then $\mathbb{E}[x] \leq \alpha + \gamma$. We assume all logarithms are natural unless otherwise specified, and we use $\mathbf{1}$ to denote the vector of all ones.

2.1. Families of Local Dynamics

We begin by describing several families of *local dynamics* for the decentralized bandit setting. We call the first family *adoption dynamics*, defined from the perspective of any node u at round t:

Adoption Dynamics: given a non-decreasing *adoption function* $f : \mathbb{R} \to [0, 1]$, for each $u \in [n]$:

(i) *At round t, assume*: node u adopted arm j; u sampled node v; and v adopted arm $k \in [m]$.
(ii) *At round $t+1$*: node u adopts arm k with probability $f(g_k^t)$ and arm j otherwise.

Note here that each node's adoption strategy at round $t+1$ depends only on the reward g_k^t obtained by its neighbor. A natural choice for f is the sigmoid function with parameter β, and we call the resulting protocol β-sigmoid-adopt. Now consider the sequence of distributions $\{\mathbf{p}^t\}$ induced by running an adoption dynamics with reward sequence $\{\mathbf{g}^t\}$. In conditional expectation through round t, we show that the mass of each coordinate j grows by a multiplicative factor with magnitude $1 + f(g_j^t) - \langle \mathbf{p}^t, f(\mathbf{g}^t)\rangle$. This is captured formally in Proposition A.1 in Appendix A.

Our second family of dynamics, which we call *comparison dynamics*, shares a similar property in conditional expectation. We first define this second family as follows:

Comparison Dynamics: given a non-decreasing score function $h : \mathbb{R} \to \mathbb{R}_{\geq 0}$, for each $u \in [n]$:

(i) *At round t, assume*: node u adopted arm j; u sampled node v; and v adopted arm $k \in [m]$.
(ii) *At round $t+1$*: define $\rho_j \propto h\big(g_j^t\big)$ and $\rho_k \propto h\big(g_k^t\big)$.
Then node u adopts arm j with probability ρ_j and arm k with probability ρ_k.

Here, a node considers both its own and its neighbor's most recent reward observation to determine its next decision (in particular, by comparing the "scores" of the rewards under h). As a natural example, we can instantiate a comparison dynamics with an exponential score function, which results in ρ_j and ρ_k forming a *softmax* distribution. We call the resulting protocol β-softmax-compare.

As mentioned, every instantiation of a comparison dynamics yields a coordinate-wise multiplicative update rule (in conditional expectation) similar to those of adoption dynamics. In particular, in Proposition A.2 in Appendix A, we derive an conditionally expected update rule for comparison dynamics similar to that of Proposition A.1.

Together, Propositions A.1 and A.2 show that under both families of dynamics, the new, conditionally expected adoption masses $\widehat{\mathbf{p}}_{t+1} := \mathbb{E}_t[\mathbf{p}^{t+1}]$ take on the more general form $\widehat{p}_j^{t+1} := \mathbb{E}_t[p_j^{t+1}] = p_j^t \cdot \big(1 + F_j(\mathbf{p}^t, \mathbf{g}^t)\big)$, where the set of m functions F_j satisfies $\sum_j p_j \cdot F_j(\mathbf{p}, \mathbf{g}) = 0$ for all $\mathbf{p} \in \Delta_m$ and $\mathbf{g} \in \mathbb{R}^m$. This observation motivates us to define the set of processes $\{\mathbf{q}^t\}$ that evolve according to such multiplicative updates at each step. Specifically, consider a sequence of distributions $\{\mathbf{q}^t\}$ that evolves according to the following definition:

Definition 2.1 (Zero-Sum Multiplicative Weights Update) *Let $\mathcal{F} = \{F_j\}_{j\in[m]}$ be a a family of m potential functions $F_j : \Delta_m \times \mathbb{R}^m \to [-1, 1]$ satisfying the zero-sum condition*

$$\textstyle\sum_{j\in[m]} q_j \cdot F_j(\mathbf{q}, \mathbf{g}) = 0 \tag{2}$$

for all $\mathbf{q} \in \Delta_m$ and $\mathbf{g} \in \mathbb{R}^m$. Then initialized from $\mathbf{q}^0 \in \Delta_m$ and given T, we say the sequence $\{\mathbf{q}^t\}$ is a zero-sum MWU process with reward sequence $\{\mathbf{g}^t\}$ if for all $t \in [T]$ and $j \in [m]$:

$$q_j^{t+1} = q_j^t \cdot \big(1 + F_j\big(\mathbf{q}^t, \mathbf{g}^t\big)\big) . \tag{3}$$

Compared to the standard (linear) versions of MWU methods [3], the zero-sum condition (2) always ensures the set of updated weights in (3) remains a distribution, *without* an additional renormalization step (i.e., the simplex Δ_m is invariant to $\{\mathbf{q}^t\}$). Further below, we develop general regret

bounds for these zero-sum MWU processes with respect to $\{\mathbf{g}^t\}$, where the bounds depend on some *quality* measure of the family $\mathcal{F}$ in distinguishing higher-mean and lower-mean rewards.

2.2. Analysis Framework for Bounding the Regret of the Induced Process $\{\mathbf{p}^t\}$

While the (coordinate-wise) iterates of each $\mathbf{p}^t$ induced by our dynamics have the same update form as those of a zero-sum MWU process in conditional expectation, neither the sequence $\{\mathbf{p}^t\}$ nor the sequence $\{\mathbb{E}_t[\mathbf{p}^{t+1}]\}$ is *itself* such a process (in the sense of Definition 2.1). However, to analyze the regret of $\{\mathbf{p}^t\}$, we introduce a true zero-sum MWU process $\{\mathbf{q}^t\}$ that starts at the same initial distribution, runs on the same reward sequence, and uses the same family $\mathcal{F}$ as follows:

Definition 2.2 (Coupled Trajectories) *Let $\mathcal{F} = \{F_j\}$ be a family of zero-sum functions as in Definition 2.1. Then given a reward sequence $\{\mathbf{g}^t\}$, consider the sequences of distributions $\{\mathbf{p}^t\}$, $\{\widehat{\mathbf{p}}^t\}$, and $\{\mathbf{q}^t\}$, each initialized at $\mathbf{p}^0 \in \Delta^m$, such that for all $j \in [m]$:*

$$q_j^{t+1} \;:=\; q_j^t \cdot \big(1 + F_j(\mathbf{q}^t, \mathbf{g}^t)\big)\;, \tag{4}$$

$$\widehat{p}_j^{t+1} := \mathbb{E}_t[p_j^{t+1}] \;=\; p_j^t \cdot \big(1 + F_j(\mathbf{p}^t, \mathbf{g}^t)\big)\;, \tag{5}$$

and where p_j^t is the average of n i.i.d. indicator random variables, each with conditional mean $\widehat{p}_j^t$.

Given this coupling definition, a straightforward calculation (derived in Appendix B) shows that we can approximate the regret $R(T)$ of the sequence $\{\mathbf{p}^t\}$ from expression (1) as follows:

$$R(T) \;\le\; \widehat{R}(T) \;:=\; \max_{j\in[m]} \sum_{t\in[T]} \mu_j^t \;-\; \sum_{t\in[T]} \mathbb{E}[\langle \mathbf{q}^t, \mathbf{g}^t\rangle] \;+\; \sum_{t\in[T]} \sigma \cdot \mathbb{E}\|\mathbf{p}^t - \mathbf{q}^t\|_1\;. \tag{6}$$

This (over)-approximation allows us to decompose $R(T)$ into (a) the regret of the zero-sum MWU process $\{\mathbf{q}^t\}$ (the difference of the first two terms) and (b) the error of the coupling (the final term). This can be roughly viewed as an *approximation error* and an *estimation error*, respectively. In Appendix B, we introduce a general analysis framework for bounding each of these quantities.

2.3. Regret Bounds for Local Dynamics

Using our analysis framework, we derive instantiated regret bounds for our local dynamics. However, given that the means $\boldsymbol{\mu}^t$ can be chosen adversarially, observe that if the number of rounds T can grow arbitrarily large, then $R(T)$ can grow linearly in T: for sequences $\{\mathbf{p}^t\}$ induced by our dynamics, once any p_j^t goes to 0 (which can occur with non-zero probability in each round), then this mass remains 0 for all subsequent rounds. Adversarially setting the j'th reward to be maximal could then lead to constant regret per round. For this reason, our regret bounds impose some constraints on T. In particular, in Appendix F we show that as long as T grows at most logarithimically in the number of nodes n, then starting from $\mathbf{p}^0 = \mathbf{1}/m$, every p_j^t is at least $1/n$ with high probability (i.e., at least one node adopts each arm in round t). This translates into a (pessimistic) constraint on the T for which we can state meaningful bounds. Specifically, we obtain the following (average) regret for the β-softmax-compare and β-sigmoid-adopt dynamics (proved in Appendix F):

Theorem 2.3 *Consider the sequence $\{\mathbf{p}^t\}$ induced by running the β-softmax-compare or β-sigmoid-adopt protocol on an (adversarial) reward sequence $\{\mathbf{g}^t\}$ intialized from $\mathbf{p}^0 = \mathbf{1}/m$. Then*

for any $c \geq 1$ *and* $n \geq 3c \log n$*, and assuming that* $T \leq (\frac{1}{2} - \epsilon) \log_5 n = O\big(\log\big(\frac{n}{m^2 \log n}\big)\big)$ *for some* $\epsilon \in (0, \frac{1}{2})$*, setting* $\beta := \sqrt{(\log m)/T}$ *yields average regret of:*

$$\frac{1}{T} \cdot R(T) \;\leq\; O\left(\sqrt{\frac{\log m}{T}}\right) + \widetilde{O}\left(\frac{\sigma m}{n^{\epsilon}} + \frac{\sigma m}{n^{c}}\right).$$

2.4. Application: Convex Optimization over the Simplex

As an application of our local dynamics and analysis framework (in particular, the regret bounds of Theorem 2.3 for the adversarial setting), we show how the protocols β-softmax-compare and β-sigmoid-adopt can approximately optimize convex functions $f : \Delta_m \to \mathbb{R}$ over the simplex when the reward sequence $\{\mathbf{g}^t\}$ is generated using a (stochastic) gradient oracle. In particular we assume:

Assumption 1 *Given a function* $f : \Delta_m \to \mathbb{R}$*, we assume that:*

(i) *f is convex with gradients bounded by* $\|\nabla f(\mathbf{q})\|_\infty \leq G$ *for all* $\mathbf{q} \in \Delta_m$*, for some* $G > 0$*.*

(ii) *At every round* $t \in [T]$*, the reward vector* $\mathbf{g}^t$ *is of the form:* $\mathbf{g}^t := -(\nabla f(\mathbf{p}^t)/G) + \mathbf{b}^t$*, where* $\mathbf{b}^t \in [-\sigma, \sigma]^m$ *is a coordinate-wise bounded random vector for some* $\sigma \in [1, 10]$*.*

Observe that condition (ii) ensures that the vector $\mathbf{g}^t$ satisfies the reward distribution conditions of our bandit setting (in particular, with $[-\sigma, \sigma]$-bounded support, and $[-1, 1]$-bounded means) and thus our regret bounds from Section 2.3 can be applied. To this end, by adapting standard reductions between MWU algorithms and (online) convex optimization [3, 13], we can use the more general adversarial regret bound of Theorem 2.3 to derive the following result, proved in Appendix G:

Theorem 2.4 *Given a convex function* $f : \Delta_m \to \mathbb{R}$*, consider the sequence* $\{\mathbf{p}^t\}$ *induced by running the* β-softmax-compare *or* β-sigmoid-adopt *protocol on a reward sequence* $\{\mathbf{g}^t\}$ *generated as in Assumption 1 with gradient bound* G*. Then for any* $c \geq 1$ *and* $n \geq 3c \log n$*, assume that* $T \leq (\frac{1}{2} - \epsilon) \log_5 n = O\big(\log\big(\frac{n}{m^2 \log n}\big)\big)$ *for some* $\epsilon \in (0, \frac{1}{2})$*, and set* $\beta := \sqrt{(\log m)/T}$*. Let* $\widetilde{\mathbf{p}} := \frac{1}{T} \sum_{t \in [T]} \mathbf{p}^t$ *denote the average arm distribution over* T *rounds. Then:*

$$\mathbb{E}[f(\widetilde{\mathbf{p}})] - \min_{\mathbf{p} \in \Delta_m} f(\mathbf{p}) \;\leq\; O\left(\sqrt{\frac{G^2 \log m}{T}}\right) + \widetilde{O}\left(G \cdot \left(\frac{\sigma m}{n^{\epsilon}} + \frac{\sigma m}{n^{c}}\right)\right).$$

3. Conclusion

To conclude, we introduced several families of dynamics for the *decentralized* bandit problem, whose regret in an adversarial reward setting grows sublinearly over rounds (so long as the total number of rounds is at most logarithmic in the size of the population). In particular, note that relative to prior related work [10, 25, 30], these are the first such dynamics that can tolerate rewards whose means are non-stationary. As an application, we showed how these dynamics can approximately optimize convex functions over the simplex at a population level. It remains open to establish optimal regret bounds in the general adversarial reward setting, and to also establish tighter bounds for longer time horizons when the reward sequences have additional (non-adversarial) structure. Moreover, analyzing these dynamics over a non-complete communication graph (in particular, understanding how mixing properties of the graph affect regret) is left as future work.

References

[1] Benjamin Allen and Martin A Nowak. Games on graphs. *EMS surveys in mathematical sciences*, 1(1):113–151, 2014.

[2] Animashree Anandkumar, Nithin Michael, Ao Kevin Tang, and Ananthram Swami. Distributed algorithms for learning and cognitive medium access with logarithmic regret. *IEEE Journal on Selected Areas in Communications*, 29(4):731–745, 2011.

[3] Sanjeev Arora, Elad Hazan, and Satyen Kale. The multiplicative weights update method: a meta-algorithm and applications. *Theory Comput.*, 8(1):121–164, 2012. doi: 10.4086/toc.2012.v008a006. URL https://doi.org/10.4086/toc.2012.v008a006.

[4] Orly Avner and Shie Mannor. Concurrent bandits and cognitive radio networks. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2014, Nancy, France, September 15-19, 2014. Proceedings, Part I 14*, pages 66–81. Springer, 2014.

[5] Luca Becchetti, Andrea Clementi, and Emanuele Natale. Consensus dynamics: An overview. *ACM SIGACT News*, 51(1):58–104, 2020.

[6] Etienne Boursier and Vianney Perchet. Sic-mmab: synchronisation involves communication in multiplayer multi-armed bandits. *Advances in Neural Information Processing Systems*, 32, 2019.

[7] Stephen Boyd, Arpita Ghosh, Balaji Prabhakar, and Devavrat Shah. Randomized gossip algorithms. *IEEE transactions on information theory*, 52(6):2508–2530, 2006.

[8] Sébastien Bubeck and Thomas Budzinski. Coordination without communication: optimal regret in two players multi-armed bandits. In *Conference on Learning Theory*, pages 916–939. PMLR, 2020.

[9] Sébastien Bubeck, Nicolo Cesa-Bianchi, et al. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends® in Machine Learning*, 5(1):1–122, 2012.

[10] L. Elisa Celis, Peter M. Krafft, and Nisheeth K. Vishnoi. A distributed learning dynamics in social groups. In *PODC*, pages 441–450. ACM, 2017.

[11] Elad Hazan et al. Introduction to online convex optimization. *Foundations and Trends® in Optimization*, 2(3-4):157–325, 2016.

[12] Eshcar Hillel, Zohar S Karnin, Tomer Koren, Ronny Lempel, and Oren Somekh. Distributed exploration in multi-armed bandits. *Advances in Neural Information Processing Systems*, 26, 2013.

[13] Jyrki Kivinen and Manfred K Warmuth. Exponentiated gradient versus gradient descent for linear predictors. *information and computation*, 132(1):1–63, 1997.

[14] Ravi Kumar Kolla, Krishna Jagannathan, and Aditya Gopalan. Collaborative learning of stochastic bandits over a social network. *IEEE/ACM Transactions on Networking*, 26(4): 1782–1795, 2018.

[15] Anastasia Koloskova, Tao Lin, Sebastian U Stich, and Martin Jaggi. Decentralized deep learning with arbitrary communication compression. *arXiv preprint arXiv:1907.09356*, 2019.

[16] Anastasia Koloskova, Sebastian U. Stich, and Martin Jaggi. Decentralized stochastic optimization and gossip algorithms with compressed communication. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, volume 97 of *Proceedings of Machine Learning Research*, pages 3478–3487. PMLR, 2019. URL http://proceedings.mlr.press/v97/koloskova19a.html.

[17] Anastasia Koloskova, Nicolas Loizou, Sadra Boreiri, Martin Jaggi, and Sebastian U Stich. A unified theory of decentralized sgd with changing topology and local updates. *arXiv preprint arXiv:2003.10422*, 2020.

[18] Lifeng Lai, Hai Jiang, and H Vincent Poor. Medium access in cognitive radio networks: A competitive multi-armed bandit framework. In *2008 42nd Asilomar Conference on Signals, Systems and Computers*, pages 98–102. IEEE, 2008.

[19] Peter Landgren, Vaibhav Srivastava, and Naomi Ehrich Leonard. Distributed cooperative decision-making in multiarmed bandits: Frequentist and bayesian algorithms. In *2016 IEEE 55th Conference on Decision and Control (CDC)*, pages 167–172. IEEE, 2016.

[20] Tor Lattimore and Csaba Szepesvári. *Bandit Algorithms*. Cambridge University Press, 2020.

[21] Xiangru Lian, Ce Zhang, Huan Zhang, Cho-Jui Hsieh, Wei Zhang, and Ji Liu. Can decentralized algorithms outperform centralized algorithms? a case study for decentralized parallel stochastic gradient descent. *Advances in neural information processing systems*, 30, 2017.

[22] Keqin Liu and Qing Zhao. Distributed learning in multi-armed bandit with multiple players. *IEEE transactions on signal processing*, 58(11):5667–5681, 2010.

[23] David Martínez-Rubio, Varun Kanade, and Patrick Rebeschini. Decentralized cooperative stochastic bandits. *Advances in Neural Information Processing Systems*, 32, 2019.

[24] Michael Mitzenmacher and Eli Upfal. *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*. Cambridge University Press, 2005. ISBN 978-0-521-83540-4. doi: 10.1017/CBO9780511813603. URL https://doi.org/10.1017/CBO9780511813603.

[25] Abishek Sankararaman, Ayalvadi Ganesh, and Sanjay Shakkottai. Social learning in multi agent multi armed bandits. *Proc. ACM Meas. Anal. Comput. Syst.*, 3(3):53:1–53:35, 2019. doi: 10.1145/3366701. URL https://doi.org/10.1145/3366701.

[26] Kevin Scaman, Francis R. Bach, Sébastien Bubeck, Yin Tat Lee, and Laurent Massoulié. Optimal algorithms for smooth and strongly convex distributed optimization in networks. In Doina Precup and Yee Whye Teh, editors, *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017*, volume 70 of *Proceedings of Machine Learning Research*, pages 3027–3036. PMLR, 2017. URL http://proceedings.mlr.press/v70/scaman17a.html.

[27] Laura Schmid, Krishnendu Chatterjee, and Stefan Schmid. The evolutionary price of anarchy: Locally bounded agents in a dynamic virus game. In *23rd International Conference on Principles of Distributed Systems (OPODIS 2019)*, volume 153, page 21. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2019.

[28] Devavrat Shah et al. Gossip algorithms. *Foundations and Trends® in Networking*, 3(1):1–125, 2009.

[29] Aleksandrs Slivkins. Introduction to multi-armed bandits. *Foundations and Trends® in Machine Learning*, 12(1-2):1–286, 2019.

[30] Lili Su, Martin Zubeldia, and Nancy A. Lynch. Collaboratively learning the best option on graphs, using bounded local memory. *Proc. ACM Meas. Anal. Comput. Syst.*, 3(1):11:1–11:32, 2019. doi: 10.1145/3322205.3311082. URL https://doi.org/10.1145/3322205.3311082.

[31] Balazs Szorenyi, Róbert Busa-Fekete, István Hegedus, Róbert Ormándi, Márk Jelasity, and Balázs Kégl. Gossip-based distributed stochastic bandit algorithms. In *International conference on machine learning*, pages 19–27. PMLR, 2013.

[32] Hanlin Tang, Shaoduo Gan, Ce Zhang, Tong Zhang, and Ji Liu. Communication compression for decentralized training. In Samy Bengio, Hanna M. Wallach, Hugo Larochelle, Kristen Grauman, Nicolò Cesa-Bianchi, and Roman Garnett, editors, *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada*, pages 7663–7673, 2018. URL https://proceedings.neurips.cc/paper/2018/hash/44feb0096faa8326192570788b38c1d1-Abstract.html.

[33] Jianyu Wang and Gauri Joshi. Cooperative sgd: A unified framework for the design and analysis of communication-efficient sgd algorithms. *arXiv preprint arXiv:1808.07576*, 2018.

[34] Jingxuan Zhu, Alec Koppel, Alvaro Velasquez, and Ji Liu. Byzantine-resilient decentralized multi-armed bandits. *arXiv preprint arXiv:2310.07320*, 2023.

Appendix A. Details on Evolution of Local Dynamics in Conditional Expectation

In this appendix, we derive the conditionally expected evolution of the adoption and comparison dynamics introduced in Section 2.

A.1. Evolution of Adoption Dyanmics

For adoption dynamics, we prove the following conditionally expected update rule:

Proposition A.1 *Let $\{\mathbf{p}^t\}$ be the sequence induced by running any adoption dynamics with adoption function f and reward sequence $\{\mathbf{g}^t\}$. Then $\mathbb{E}_t[p_j^{t+1}] = p_j^t \cdot \big(1 + f(g_j^t) - \langle \mathbf{p}^t, f(\mathbf{g}^t)\rangle\big)$, for every t and $j \in [m]$, where $f(\mathbf{g}^t) \in [0,1]^m$ denotes the coordinate-wise application of f on $\mathbf{g}^t$.*

Proof First, letting $c_u^{t+1} \in [m]$ denote the arm adopted by node $u \in [n]$ in round $t+1$, observe that

$$\mathbb{E}_t[p_j^{t+1}] \;=\; \frac{1}{n} \sum_{u \in [n]} \mathbb{P}_t[c_u^{t+1} = j] \;,$$

which follows from the fact that p_j^{t+1} is the average of the n indicators $\mathbf{1}\{c_u^{t+1} = j\}$. By the local symmetry of the dynamics, $\mathbb{P}_t[c_u^{t+1} = j]$ is equal for all nodes u. However, this value is dependent on c_u^t (i.e., the adoption decision of u at the previous round t).

Thus using the law of total probability, for any node u we can write

$$\begin{aligned}\mathbb{P}_t[c_u^{t+1} = j] \;&=\; \mathbf{1}\{c_u^t = j\} \cdot \mathbb{P}_t[c_u^{t+1} = j | c_u^t = j] \\ &\quad + \sum_{k \neq j \in [m]} \mathbf{1}\{c_u^t = k\} \cdot \mathbb{P}_t[c_u^{t+1} = j | c_u^t = k] \;.\end{aligned}$$

Now fix node u, and let $v \in [n]$ denote the node the u samples in round t. Now recall from the definition of the dynamics that if $c_u^t = k \neq j$, then $c_u^{t+1} = j$ with probability $f(g_j^t)$ only if node v adopted arm j in round t, i.e., $c_v^t = j$. On the other hand, if $c_u^t = j$, then $c_u^{t+1} = j$ either if $c_v^t = j$, or if $c_v^t = k \neq j$ and node u *rejects* adopting arm k with probability $1 - f(g_k^t)$.

Thus we have

$$\begin{aligned}\mathbb{P}_t[c_u^{t+1} = j | c_u^t = j] \;&=\; p_j^t + \sum_{k \neq j \in [m]} p_k^t \cdot (1 - f(g_k^t)) \\ \text{and} \quad \mathbb{P}_t[c_u^{t+1} = j | c_u^t = k] \;&=\; p_j^t \cdot f(g_j^t) \quad \text{for } k \neq j \;.\end{aligned}$$

Combining these cases, noting also that $\frac{1}{n}\sum_{u \in [v]} \mathbf{1}\{c_u^t = k\} = p_k^t$ for any $k \in [m]$, and using the fact that $\sum_{k \in [m]} p_k^t = 1$, we can then write

$$\begin{aligned}\mathbb{E}_t[p_j^{t+1}] \;&=\; p_j^t \cdot \big(p_j^t + \sum_{k \neq j \in [m]} p_k^t \cdot (1 - f(g_k^t))\big) + \sum_{k \neq j \in [m]} p_k^t \cdot \big(p_j^t \cdot f(g_j^t)\big) \\ &=\; p_j^t \cdot \Big(1 + \sum_{k \neq j \in [m]} p_k^t \cdot \big(f(g_j^t) - f(g_k^t)\big)\Big) \\ &=\; p_j^t \cdot \Big(1 + f(g_j^t) - \big\langle \mathbf{p}^t, f(\mathbf{g}^t) \big\rangle\Big) \;,\end{aligned}$$

which concludes the proof. ■

Importantly, we also verify that such multiplicative updates in every coordinate j still lead to a proper distribution: for this, it is easy to check that

$$\begin{aligned}\sum_{j \in [m]} \mathbb{E}_t[p_j^{t+1}] \;&=\; \sum_{j \in [m]} p_j^t \cdot \Big(1 + f(g_j^t) - \big\langle \mathbf{p}^t, f(\mathbf{g}^t) \big\rangle\Big) \\ &=\; \sum_{j \in [m]} p_j^t + \big\langle \mathbf{p}^t, f(\mathbf{g}^t) \big\rangle - \big\langle \mathbf{p}^t, f(\mathbf{g}^t) \big\rangle \;=\; 1 \;,\end{aligned}$$

which holds since $\mathbf{p}^t$ is a distribution.

Finally, recall from Section 2.1 that when the adoption function f is a sigmoid function with parameter β we call the resulting protocol β-sigmoid-adopt. Stated formally:

Local Protocol 1 (β-sigmoid-adoption) *Let* β-sigmoid-adopt *be the adoption dynamics protocol instantiated by the adoption function* $f_\beta(g) := \frac{1}{1+\exp(-\beta \cdot g)}$ *for* $\beta \in [0,1]$ *and any* $g \in \mathbb{R}$.

A.2. Evolution of Comparison Dynamics

For comparison dynamics, we develop an update rule in conditional expectation whose form is analogous to that of Proposition A.1. Specifically, we show:

Proposition A.2 *Let* $\{\mathbf{p}^t\}$ *be the sequence induced by running any comparison dynamics with score function* h *and reward seqeunce* $\{\mathbf{g}^t\}$. *Furthermore, for any* $\mathbf{g} \in \mathbb{R}^m$ *and* $j \in [m]$, *let* $H(\mathbf{g}, j) \in [-1,1]^m$ *be the* m*-dimensional vector whose* k*'th coordinate is given by* $\rho_j - \rho_k$. *Then*

$$\mathbb{E}_t[p_j^{t+1}] = p_j^t \cdot \big(1 + \langle \mathbf{p}^t, H(\mathbf{g}^t, j)\rangle\big)$$

for every $t \in [T]$ *and* $j \in [m]$.

Proof Fix $j \in [m]$ and $t \in [T]$. Again let $c_i^t \in [m]$ denote the arm adopted by node $i \in [n]$ at round t. Then observe that we can write

$$\mathbb{E}_t\big[p_j^{t+1}\big] \;=\; \frac{1}{n}\sum_{i\in[n]} \mathbb{P}_t\big[c_i^{t+1} = j\big] \;=\; \frac{1}{n}\sum_{i\in[n]} \Bigg(\mathbf{1}\{c_i^t = j\} \cdot \mathbb{P}_t\big[c_i^{t+1} = j \mid c_i^t = j\big] + \sum_{k\neq j\in[m]} \mathbf{1}\{c_i^t = k\} \cdot \mathbb{P}_t\big[c_i^{t+1} = j \mid c_i^t = k\big] \Bigg) .$$

In the case that $c_i^t = j$, note that $c^{t+1} = j$ with probability 1 if node i samples a neighbor u that also pulled arm j in round t. Otherwise, if u pulled some arm $k \neq j$, then node i adopts j with probability $1 - h(g_k^t)/\big(h(g_j^t) + h(g_k^t)\big)$. Together, this means that

$$\begin{aligned} \mathbb{P}_t\Big[c_i^{t+1} = j \mid c_i^t = j\Big] \;&=\; p_j^t + \sum_{k\neq j\in[m]} p_k^t \left(1 - \frac{h(g_k^t)}{h(g_j^t) + h(g_k^t)}\right) \\ &=\; 1 - \sum_{k\neq j\in[m]} p_k^t \cdot \frac{h(g_k^t)}{h(g_j^t) + h(g_k^t)} \quad . \end{aligned} \tag{7}$$

In the other case when $c_i^t = k \neq j$, then $c^{t+1} = j$ only when node i samples a neighbor that pulled arm j in round t, and thus

$$\mathbb{P}_t\big[c_i^{t+1} = j \mid c_i^t = k\big] \;=\; p_j^t \cdot \frac{h(g_j^t)}{h(g_j^t) + h(g_k^t)} \,. \tag{8}$$

Now observe that for any $k \in [m]$, $\frac{1}{n}\sum_{i\in[n]} \mathbf{1}\{c_i^t = k\} = p_k^t$ by definition. Then together with expression (7) and (8), we have

$$\begin{aligned}
\mathbb{E}_t\left[p_j^{t+1}\right] &= p_j^t\left(1 - \sum_{k\neq j\in[m]} p_k^t \cdot \frac{h(g_k^t)}{h(g_j^t)+h(g_k^t)}\right) + \sum_{k\neq j\in[m]} p_k^t \cdot p_j^t \frac{h(g_j^t)}{h(g_j^t)+h(g_k^t)} \\
&= p_j^t \cdot \left[1 + \sum_{k\neq j\in[m]} p_k^t \cdot \frac{h(g_j^t)-h(g_k^t)}{h(g_j^t)+h(g_k^t)}\right] \\
&= p_j^t \cdot \left[1 + \left\langle \mathbf{p}^t, H(\mathbf{g}^t, j)\right\rangle\right],
\end{aligned}$$

which concludes the proof. ■

Again, we also verify that for any $\mathbf{p} \in \Delta_m$ and $\mathbf{g} \in [0,1]^m$, the family of functions $\{\langle \mathbf{p}, H(\mathbf{g}, j)\rangle\}_{j\in[m]}$ satisfies the zero-sum property $\sum_{j\in[m]} p_j \cdot \langle \mathbf{p}, H(\mathbf{g}, j)\rangle = 0$. To see this, observe that

$$\begin{aligned}
\sum_{j\in[m]} p_j \cdot \langle \mathbf{p}, H(\mathbf{g}, j)\rangle &= \sum_{j\in[m]} p_j \cdot \sum_{k\in[m]} p_k \cdot \frac{h(g_j^t)-h(g_k^t)}{h(g_j^t)+h(g_k^t)} \\
&= \sum_{(j,k)\in[m]\times[m]} p_j \cdot p_k \cdot \left(\frac{h(g_j^t)-h(g_k^t)}{h(g_j^t)+h(g_k^t)} + \frac{h(g_k^t)-h(g_j^t)}{h(g_j^t)+h(g_k^t)}\right) = 0\,.
\end{aligned}$$

Finally, recall that when the score function h is an exponential with parameter β, we call the resulting protocol β-softmax-compare. Defined formally:

Local Protocol 2 (β-softmax-comparison) *Let* β-softmax-compare *denote the comparison dynamics protocol instantiated with the score function* $h_\beta(g) := \exp(\beta \cdot g)$ *for some* $\beta \in [0,1]$.

Appendix B. Details on Analysis Framework

In this section, we provide more details on the analysis framework from Section 2.2, and specifically on approximating the regret $R(T)$ in the context of the coupling from Definition 2.2. For convenenience, we first restate this coupling definition:

Definition 2.2 (Coupled Trajectories) *Let* $\mathcal{F} = \{F_j\}$ *be a family of zero-sum functions as in Definition 2.1. Then given a reward sequence* $\{\mathbf{g}^t\}$*, consider the sequences of distributions* $\{\mathbf{p}^t\}$*,* $\{\widehat{\mathbf{p}}^t\}$*, and* $\{\mathbf{q}^t\}$*, each initialized at* $\mathbf{p}^0 \in \Delta^m$*, such that for all* $j \in [m]$*:*

$$q_j^{t+1} := q_j^t \cdot \left(1 + F_j(\mathbf{q}^t, \mathbf{g}^t)\right), \tag{4}$$

$$\widehat{p}_j^{t+1} := \mathbb{E}_t[p_j^{t+1}] = p_j^t \cdot \left(1 + F_j(\mathbf{p}^t, \mathbf{g}^t)\right), \tag{5}$$

and where p_j^t *is the average of* n *i.i.d. indicator random variables, each with conditional mean* $\widehat{p}_j^t$*.*

Now recall from expressions (1) and (6) that we define

$$R(T) \;:=\; \max_{j\in[m]} \sum_{t\in[T]} \mu_j^t - \sum_{t\in[T]} \mathbb{E}[\langle \mathbf{p}^t, \mathbf{g}^t\rangle]$$
$$\text{and} \quad \widehat{R}(T) \;:=\; \max_{j\in[m]} \sum_{t\in[T]} \mu_j^t - \sum_{t\in[T]} \mathbb{E}[\langle \mathbf{q}^t, \mathbf{g}^t\rangle] + \sum_{t\in[T]} \sigma \cdot \mathbb{E}\|\mathbf{p}^t - \mathbf{q}^t\|_1 \;,$$

where we assume $\{\mathbf{p}^t\}$ and $\{\mathbf{q}^t\}$ are specified by the coupling from Definition 2.2. Here, we show that $R(T) \;\le\; \widehat{R}(T)$, which was stated without proof in expression (6) in Section 2.2.

Proposition B.1 *For any sequences $\{\mathbf{p}^t\}$ and $\{\mathbf{q}^t\}$ as in Definition 2.2. Then $R(T) \le \widehat{R}(T)$ with respect to any reward sequence $\{\mathbf{g}^t\}$ where each $\mathbf{g}^t \in [-\sigma, \sigma]^m$.*

Proof First, observe that for every $t \in [T]$, we can write

$$\mathbb{E}[\langle \mathbf{p}^t, \mathbf{g}^t\rangle] \;=\; \mathbb{E}[\langle \mathbf{q}^t, \mathbf{g}^t\rangle] - \mathbb{E}[\langle \mathbf{q}^t - \mathbf{p}^t, \mathbf{g}^t\rangle] \;.$$

Now recall from the definition of the problem setting that the randomness of $\mathbf{g}^t$ is independent from that of both $\mathbf{q}^t$ and $\mathbf{p}^t$. Thus together with Hölder's inequality, it follows that

$$\begin{aligned} R(T) \;&=\; \max_{j\in[m]} \sum_{t\in[T]} \mu_j^t - \sum_{t\in[T]} \mathbb{E}[\langle \mathbf{q}^t, \mathbf{g}^t\rangle] + \sum_{t\in[T]} \mathbb{E}[\langle \mathbf{q}^t - \mathbf{p}^t, \mathbf{g}^t\rangle] \\ &\le\; \max_{j\in[m]} \sum_{t\in[T]} \mu_j^t - \sum_{t\in[T]} \mathbb{E}[\langle \mathbf{q}^t, \mathbf{g}^t\rangle] + \sum_{t\in[T]} \mathbb{E}\big[\|\mathbf{q}^t - \mathbf{p}^t\|_1 \cdot \|\mathbf{g}^t\|_\infty\big] \;. \end{aligned}$$

Now by the assumption that for each $t \in [T]$, every coordinate g_j^t of $\mathbf{g}^t$ is drawn from a distribution whose support is bounded in $[-\sigma, \sigma]$, we have $\|\mathbf{g}^t\|_\infty \le \sigma$. Thus we conclude

$$R(T) \;\le\; \max_{j\in[m]} \sum_{t\in[T]} \mu_j^t - \sum_{t\in[T]} \mathbb{E}[\langle \mathbf{q}^t, \mathbf{g}^t\rangle] + \sum_{t\in[T]} \sigma \cdot \mathbb{E}\|\mathbf{q}^t - \mathbf{p}^t\|_1 \;=:\; \widehat{R}(T) \;,$$

which conludes the proof. ∎

B.1. Overview of Analysis Framework

We now give an overview of our framework for bounding the quantity $\widehat{R}(T)$. Our approach has two steps: first, bounding the regret of the zero-sum MWU process (i.e., the difference of the first two summations of $\widehat{R}(T)$), and second, bounding the error terms induced by the coupling (i.e., the third summation in $\widehat{R}(T)$).

Regret Bounds for the Zero-Sum MWU Process To analyze the regret of the zero-sum MWU process using a family $\mathcal{F}$, the key step is to relate the the function value $F_j(\mathbf{q}, \mathbf{g})$ in conditional expectation to the difference $\mu_j - \langle \mathbf{q}, \boldsymbol{\mu}^t\rangle$, which measures the relative magnitude of the j'th arm's mean to the globally-weighted average. To this end we make the following assumptions on $\mathcal{F}$:

Assumption 2 *Let $\mathcal{F} = \{F_j\}$ be a family of potential functions satisfying the zero-sum condition from Definition 2.1, and let $\{\mathbf{g}^t\}$ be a sequence of rewards. Then we assume there exist constants $0 < \alpha_1 \le \alpha_2 < 1/4$, $\delta \in [0, 1]$, and $L > 0$ such that for all j and $\mathbf{g}^t$:*

(i) for all $\mathbf{q} \in \Delta_m$: $\frac{\alpha_1}{3}\left|\mu_j^t - \langle \mathbf{q}, \boldsymbol{\mu}^t \rangle - \delta\right| \le \left|\mathbb{E}_{\mathbf{q}}\left[F_j(\mathbf{q}, \mathbf{g}^t)\right]\right| \le \frac{\alpha_2}{3}\left|\mu_j^t - \langle \mathbf{q}, \boldsymbol{\mu}^t \rangle + \delta\right|$

(ii) for all $\mathbf{p}, \mathbf{q} \in \Delta_m$: $|F_j(\mathbf{q}, \mathbf{g}^t) - F_j(\mathbf{p}, \mathbf{g}^t)| \le L \cdot \|\mathbf{p} - \mathbf{q}\|_1$.

Under this assumption, we prove (in Appendix C) the following bound on the expected regret of a zero-sum MWU process, which is parameterized with respect to constants α_1, α_2, and δ:

Theorem B.2 *Consider a $T \ge 1$ round zero-sum MWU process $\{\mathbf{q}^t\}$ from Definition 2.1 with reward sequence $\{\mathbf{g}^t\}$ and using a family $\mathcal{F}$ that satisfies Assumption 2 with parameters α_1, α_2 and δ, and assume that $q_j^0 \ge \rho > 0$. Then for every $j \in [m]$:*

$$\sum_{t\in[T]} \mu_j^t - \sum_{t\in[T]} \mathbb{E}[\langle \mathbf{q}^t, \mathbf{g}^t \rangle] \le \frac{3\log(1/\rho)}{\alpha_1} + 2\left(\frac{\alpha_2^2}{\alpha_1} + \frac{\alpha_2 - \alpha_1}{\alpha_1} + \frac{\delta\alpha_2}{\alpha_1}\right) \cdot T \,.$$

Intuitively, condition (i) of Assumption 2 specifies a two-sided *multiplicative* correlation between $\mathbb{E}_{\mathbf{q}}[F_j(\mathbf{q}, \mathbf{g}^t)]$ and $\mu_j^t - \langle \mathbf{q}, \boldsymbol{\mu}^t \rangle$, while also allowing for some additive slack δ. The sharpness of Theorem B.2 depends on the tightness of this correlation. In particular, if $\alpha_1 = \alpha_2 = \alpha$ and $\delta = 0$, and supposing that $q_j^0 = 1/m$ deterministically, then the right hand side in the theorem recovers the standard (and optimal) MWU regret bounds [3]. We thus generally wish that the family $\mathcal{F}$ induced by our local dynamics satisfies condition (i) of the assumption with $\alpha_1 = \alpha_2$, and $\delta = O(\alpha_1)$, where α_1 has some dependence on a free, tunable parameter. For the β-sigmoid-adopt and β-softmax-compare protocols introduced earlier, we show this is exactly the case, and we provide the formal statements and proofs in Appendix D.

Controlling the Coupling Error The second step in the general analysis framework is to bound the error $\sum_{t\in[T]} \mathbb{E}\|\mathbf{p}^t - \mathbf{q}^t\|_1$ of the coupled trajectories. For this, we control the growth of each term in the sum by leveraging property (ii) of Assumption 2, and by applying standard concentration bounds. In Appendix E, we describe this approach in more details and prove the following lemma:

Lemma B.3 *Consider the sequences $\{\mathbf{p}^t\}$, $\{\widehat{\mathbf{p}}^t\}$, and $\{\mathbf{q}^t\}$ from Definition 2.2 with a reward sequence $\{\mathbf{g}^t\}$ and using a family $\mathcal{F}$ that satisfies Assumption 2 with parameter L. Let $\kappa := (3+L)$, and assume $n \ge 3c\log n$ for some $c \ge 1$. Then for any $T \ge 1$:*

$$\sum_{t\in[T]} \mathbb{E}\|\mathbf{q}^{t+1} - \mathbf{p}^{t+1}\|_1 \le \widetilde{O}\left(\frac{m \cdot \kappa^T}{\sqrt{n}} + \frac{m \cdot T}{n^c}\right) .$$

Appendix C. Details on Zero-Sum Multiplicative Weight Updates

In this section, we prove the regret bound on the zero-sum MWU process from Theorem B.2, which is restated here:

Theorem B.2 *Consider a $T \ge 1$ round zero-sum MWU process $\{\mathbf{q}^t\}$ from Definition 2.1 with reward sequence $\{\mathbf{g}^t\}$ and using a family $\mathcal{F}$ that satisfies Assumption 2 with parameters α_1, α_2 and δ, and assume that $q_j^0 \ge \rho > 0$. Then for every $j \in [m]$:*

$$\sum_{t\in[T]} \mu_j^t - \sum_{t\in[T]} \mathbb{E}[\langle \mathbf{q}^t, \mathbf{g}^t \rangle] \le \frac{3\log(1/\rho)}{\alpha_1} + 2\left(\frac{\alpha_2^2}{\alpha_1} + \frac{\alpha_2 - \alpha_1}{\alpha_1} + \frac{\delta\alpha_2}{\alpha_1}\right) \cdot T \,.$$

For convenience, we also restate Assumption 2:

Assumption 2 *Let $\mathcal{F} = \{F_j\}$ be a family of potential functions satisfying the zero-sum condition from Definition 2.1, and let $\{\mathbf{g}^t\}$ be a sequence of rewards. Then we assume there exist constants $0 < \alpha_1 \leq \alpha_2 < 1/4$, $\delta \in [0, 1]$, and $L > 0$ such that for all j and $\mathbf{g}^t$:*

(i) for all $\mathbf{q} \in \Delta_m$: $\frac{\alpha_1}{3}\left|\mu_j^t - \langle \mathbf{q}, \boldsymbol{\mu}^t \rangle - \delta\right| \leq \left|\mathbb{E}_{\mathbf{q}}\left[F_j(\mathbf{q}, \mathbf{g}^t)\right]\right| \leq \frac{\alpha_2}{3}\left|\mu_j^t - \langle \mathbf{q}, \boldsymbol{\mu}^t \rangle + \delta\right|$

(ii) for all $\mathbf{p}, \mathbf{q} \in \Delta_m$: $|F_j(\mathbf{q}, \mathbf{g}^t) - F_j(\mathbf{p}, \mathbf{g}^t)| \leq L \cdot \|\mathbf{p} - \mathbf{q}\|_1$.

Roughly speaking, condition (i) of the assumption allows us to relate each $F_j(\mathbf{q}, \mathbf{g})$ to g_j in (conditional) expectation. From there, we can leverage standard approaches to proving MWU regret bounds (i.e., in the spirit of Arora et al. [3]), but with some additional bookkeeping to account for the α_1, α_2, and δ parameters. We also allow for a probabilistic lower bound on the initial mass at the j'th coordinate, which is useful for deriving the epoch-based regret bounds from Section 2.3.

Proof *(of Theorem B.2)* Fix $j \in [m]$ and $t \in [T]$. Recall that in round t, both $\mathbf{q}^t$ and $\mathbf{g}^t$ are random variables, where $\mathbf{q}^t$ depends on the randomness in both $\{\mathbf{q}^{t-1}\}$ and $\{\mathbf{g}^{t-1}\}$. Then conditioning on both of these sequences (which is captured in the notation $\mathbb{E}_{t-1}[\cdot]$), we can use the definition of the update rule in expression (3) to write

$$\begin{aligned}\mathbb{E}_{t-1}\left[q_j^t\right] &= \mathbb{E}_{t-1}\left[\, q_j^t \cdot \left(1 + F_j(\mathbf{q}^t, \mathbf{g}^t)\right) \,\right] \\ &= q_j^t \cdot \mathbb{E}_{t-1}\left[1 + F_j(\mathbf{q}^t, \mathbf{g}^t)\right] = q_j^t \cdot \left(1 + \mathbb{E}_{t-1}\left[F_j(\mathbf{q}^t, \mathbf{g}^t)\right]\right) .\end{aligned}$$

Here, the second equality comes from the fact that $\mathbf{q}^t$ is a constant when conditioning on $\{\mathbf{q}^{t-1}\}$ and $\{\mathbf{g}^{t-1}\}$, and thus $\mathbb{E}_{t-1}[q_j^t] = q_j^t$. Now for readability, let us define

$$m_j^t := \mathbb{E}_{t-1}\left[F_j(\mathbf{q}^t, \mathbf{g}^t)\right] ,$$

and that m_j^t is deterministic (meaning $\mathbb{E}[m_j^t] = m_j^t$), since the only remaining randomness after the conditioning is with respect to $\mathbf{g}^t$. Thus using the law of iterated expectation, we can ultimately write

$$\mathbb{E}\left[q_j^{t+1}\right] = \mathbb{E}\left[\, \mathbb{E}_{t-1}\left[q_j^{t+1}\right] \,\right] = \mathbb{E}\left[q_j^t\right] \cdot \left(1 + m_j^t\right) .$$

By repeating the preceding argument for each of $\mathbb{E}[q_j^{t-1}], \ldots, \mathbb{E}[q_j^1]$, and setting $T = t + 1$, we find that

$$\mathbb{E}\left[q_j^T\right] = q_j^0 \cdot \prod_{t \in [T-1]} \left(1 + m_j^t\right) . \tag{9}$$

From here, we roughly follow a standard multiplicative weights analysis: first, define the sets M_j^+ and M_j^- as

$$\begin{aligned} M_j^+ &= \{t \in [T-1] \,:\, m_j^t \geq 0\} \\ \text{and } M_j^- &= \{t \in [T-1] \,:\, m_j^t < 0\} , \end{aligned}$$

where clearly $M_j^+ \cup M_j^- = [T]$. Then we can rewrite expression (9) as

$$\mathbb{E}[q_j^T] = q_j^0 \cdot \prod_{t \in M_j^+} \left(1 + m_j^t\right) \cdot \prod_{t \in M_j^-} \left(1 + m_j^t\right) .$$

Now for each t, define $\Delta_j^t := \mu_j^t - \langle \mathbf{q}^t, \boldsymbol{\mu}^t \rangle \in [-2, 2]$. Using this notation, observe that Assumption 2 implies

$$\begin{aligned} m_j^t &\geq \frac{\alpha_1}{3} \cdot (\Delta_j^t - \delta) \quad \text{when } m_j^t \geq 0 \\ \text{and} \quad m_j^t &\geq \frac{\alpha_2}{3} \cdot (\Delta_j^t - \delta) \quad \text{when } m_j^t < 0 \, . \end{aligned}$$

Note that when $m_j^t < 0$, the latter inequality implies that $\Delta_j^t - \delta < 0$. On the other hand, when $m_j^t \geq 0$, the first inequality provides no further information on the sign of $\Delta_j^t - \delta$. Thus we define the additional two sets G_j^+ and G_j^- as

$$\begin{aligned} G_j^+ &= \{t \in [T-1] \, : \, \Delta_j^t - \delta \geq 0\} \\ \text{and} \quad G_j^- &= \{t \in [T-1] \, : \, \Delta_j^t - \delta < 0\} \, . \end{aligned}$$

Combining the pieces above, it follows that

$$\begin{aligned} \mathbb{E}[q_j^T] &\geq q_j^0 \cdot \prod_{t \in M_j^+} \left(1 + \alpha_1 \tfrac{(\Delta_j^t - \delta)}{3}\right) \cdot \prod_{t \in M_j^-} \left(1 + \alpha_2 \tfrac{(\Delta_j^t - \delta)}{3}\right) \\ &= q_j^0 \cdot \prod_{t \in M_j^+ \cap G_j^+} \left(1 + \alpha_1 \tfrac{(\Delta_j^t - \delta)}{3}\right) \cdot \prod_{t \in M_j^+ \cap G_j^-} \left(1 + \alpha_1 \tfrac{(\Delta_j^t - \delta)}{3}\right) \cdot \prod_{t \in M_j^-} \left(1 + \alpha_2 \tfrac{(\Delta_j^t - \delta)}{3}\right) . \end{aligned}$$

Observe also that each $|\Delta_j^t - \delta| \in [-3, 3]$ by the definition of Δ_j^t and by the assumption that $\delta \in [0, 1]$. Thus we can then use the facts that $(1 + \alpha x) \geq (1 + \alpha)^x$ for $x \in [0, 1]$, and that $(1 + \alpha x) \geq (1 - \alpha)^{-x}$ for $x \in [-1, 0]$, which allows us to further simplify and write

$$\mathbb{E}[q_j^T] \geq q_j^0 \cdot \prod_{t \in M_j^+ \cap G_j^+} (1 + \alpha_1)^{\frac{(\Delta_j^t - \delta)}{3}} \cdot \prod_{t \in M_j^+ \cap G_j^-} (1 - \alpha_1)^{-\frac{(\Delta_j^t - \delta)}{3}} \cdot \prod_{t \in M_j^-} (1 - \alpha_2)^{-\frac{(\Delta_j^t - \delta)}{3}} \, .$$

Now using the fact that $q_j^T \leq 1$, taking logarithms, and multiplying through by 3, we find

$$\begin{aligned} 0 \geq 3 \log q_j^0 &+ \sum_{t \in M_j^+ \cap G_j^+} \log(1 + \alpha_1)(\Delta_j^t - \delta) \\ &- \sum_{t \in M_j^+ \cap G_j^-} \log(1 - \alpha_1)(\Delta_j^t - \delta) \; - \sum_{t \in M_j^-} \log(1 - \alpha_2)(\Delta_j^t - \delta) \, . \end{aligned} \tag{10}$$

By definition, recall that $(\delta_j^t - \delta)$ is non-negative for $t \in M_j^+ \cap G_j^+$, and negative for $t \in M_j^+ \cap G_j^-$ and $t \in M_j^-$. Thus using the identities $\log(1 + x) \geq x - x^2$ and $-\log(1 - x) \leq x + x^2$, which

both hold for all $x \in [0, \frac{1}{2}]$, we can further bound expression (10) and rearrange to find

$$\begin{aligned}
&3\log\big(1/q_j^0\big) \\
&\quad\geq \sum_{t\in M_j^+\cap G_j^+}(\alpha_1-\alpha_1^2)(\Delta_j^t-\delta)+\sum_{t\in M_j^+\cap G_j^-}(\alpha_1+\alpha_1^2)(\Delta_j^t-\delta)+\sum_{t\in M_j^-}(\alpha_2+\alpha_2^2)(\Delta_j^t-\delta) \\
&\quad= \sum_{t\in M_j^+\cap G_j^+}\Delta_j^t\alpha_1 \;-\sum_{t\in M_j^+\cap G_j^+}\Delta_j^t\alpha_1^2 \;-\sum_{t\in M_j^+\cap G_j^+}\delta(\alpha_1-\alpha_1^2) \\
&\qquad+\sum_{t\in M_j^+\cap G_j^-}\Delta_j^t\alpha_1 \;+\sum_{t\in M_j^+\cap G_j^-}\Delta_j^t\alpha_1^2 \;-\sum_{t\in M_j^+\cap G_j^-}\delta(\alpha_1+\alpha_1^2) \\
&\qquad+\sum_{t\in M_j^-}\Delta_j^t\alpha_1 \;+\sum_{t\in M_j^-}\Delta_j^t(\alpha_2-\alpha_1) \;+\sum_{t\in M_j^-}\Delta_j^t\alpha_2^2 \;-\sum_{t\in M_j^-}\delta(\alpha_2+\alpha_2^2)\,. \qquad (11)
\end{aligned}$$

Now in expression (11), there are seven summations which we collect into four groups, bound, and simplify as follows:

$$\begin{aligned}
&\text{(i)} \quad \sum_{t\in M_j^+\cap G_j^+}\Delta_j^t\alpha_1 \;+\sum_{t\in M_j^+\cap G_j^-}\Delta_j^t\alpha_1 \;+\sum_{t\in M_j^-}\Delta_j^t\alpha_1 \;=\; \alpha_1\sum_{t\in[T]}\Delta_j^t \\
&\text{(ii)} \quad -\sum_{t\in M_j^+\cap G_j^+}\Delta_j^t\alpha_1^2 \;+\sum_{t\in M_j^+\cap G_j^-}\Delta_j^t\alpha_1^2 \;+\sum_{t\in M_j^-}\Delta_j^t\alpha_2^2 \;\geq\; -\alpha_2^2\sum_{t\in[T]}|\Delta_j^t| \\
&\text{(iii)} \quad +\sum_{t\in M_j^-}\Delta_j^t(\alpha_2-\alpha_1) \;\geq\; -(\alpha_2-\alpha_1)\sum_{t\in[T]}|\Delta_j^t| \\
&\text{(iv)} \quad -\sum_{t\in M_j^+\cap G_j^+}\delta(\alpha_1-\alpha_1^2) \;-\sum_{t\in M_j^+\cap G_j^-}\delta(\alpha_1+\alpha_1^2) \;-\sum_{t\in M_j^-}\delta(\alpha_2+\alpha_2^2) \;\geq\; -2\alpha_2\sum_{t\in[T]}\delta\,.
\end{aligned}$$

In the above, we use the fact that $\alpha_1 \leq \alpha_2 \implies \alpha_1^2 \leq \alpha_2^2$ (for (ii) and (iv)) and that $\Delta_j^t \geq -|\Delta_j^t|$ for any t (for (ii), (iii), and (iv)).

Substituting these groups back into expression (11), we ultimately find that

$$\begin{aligned}
3\log\big(1/q_j^0\big) \;&\geq\; \alpha_1\sum_{t\in[T]}\Delta_j^t-\Big((\alpha_2^2+(\alpha_2-\alpha_1))\sum_{t\in[T]}|\Delta_j^t|+2\alpha_2\sum_{t\in[T]}\delta\Big) \\
&\geq\; \alpha_1\sum_{t\in[T]}\Delta_j^t-\big((\alpha_2^2+(\alpha_2-\alpha_1))+\delta\alpha_2\big)\cdot 2T\,,
\end{aligned}$$

where the final inequality comes from the fact that $|\Delta_j^t| \leq 2$ and the assumption that $\delta \leq 1$. Thus using the definition $\Delta_j^t = \mu_j^t - \langle \mathbf{q}^t, \boldsymbol{\mu}^t\rangle$, we can rearrange to write

$$\alpha_1\sum_{t\in[T]}\big(\mu_j^t-\langle\mathbf{q}^t,\boldsymbol{\mu}^t\rangle\big) \;\leq\; 3\log\big(1/q_j^0\big)+2(\alpha_2^2+(\alpha_2-\alpha_1)+\delta\alpha_2)\cdot T\,.$$

Finally, observe for each $t \in [T]$ that $\langle \mathbf{q}^t, \boldsymbol{\mu}^t\rangle = \langle \mathbf{q}^t, \mathbb{E}[\mathbf{g}^t]\rangle = \mathbb{E}_{\mathbf{q}^t}[\langle \mathbf{q}^t, \mathbf{g}^t\rangle]$, and recall also by assumption that $q_j^0 \geq \rho > 0$, for every j, with probability at least $1-\gamma$. Thus we find

$$\sum_{t\in[T]}\mu_j^t-\sum_{t\in[T]}\mathbb{E}_{\mathbf{q}^t}[\langle\mathbf{q}^t,\mathbf{g}^t\rangle] \;\leq\; \frac{3\log(1/\rho)}{\alpha_1}+2\left(\frac{\alpha_2^2}{\alpha_1}+\frac{\alpha_2-\alpha_1}{\alpha_1}+\frac{\delta\alpha_2}{\alpha_1}\right)\cdot T\,,$$

which concludes the proof. ■

Appendix D. Details on (α, δ, L) Parameters for Local Dynamics

In this section, we derive estimates of the parameter values α_1, α_2, δ, and L for our local dynamics β-softmax-compare and β-sigmoid-adopt that are needed to satisfy Assumption 2. Recall that this assumption says the following:

Assumption 2 *Let $\mathcal{F} = \{F_j\}$ be a family of potential functions satisfying the zero-sum condition from Definition 2.1, and let $\{\mathbf{g}^t\}$ be a sequence of rewards. Then we assume there exist constants $0 < \alpha_1 \leq \alpha_2 < 1/4$, $\delta \in [0, 1]$, and $L > 0$ such that for all j and $\mathbf{g}^t$:*

(i) for all $\mathbf{q} \in \Delta_m$: $\frac{\alpha_1}{3}\left|\mu_j^t - \langle \mathbf{q}, \boldsymbol{\mu}^t \rangle - \delta\right| \leq \left|\mathbb{E}_{\mathbf{q}}\left[F_j(\mathbf{q}, \mathbf{g}^t)\right]\right| \leq \frac{\alpha_2}{3}\left|\mu_j^t - \langle \mathbf{q}, \boldsymbol{\mu}^t \rangle + \delta\right|$

(ii) for all $\mathbf{p}, \mathbf{q} \in \Delta_m$: $|F_j(\mathbf{q}, \mathbf{g}^t) - F_j(\mathbf{p}, \mathbf{g}^t)| \leq L \cdot \|\mathbf{p} - \mathbf{q}\|_1$.

We derive such satisfying constants for β-softmax-compare and β-sigmoid-adopt in Sections D.1 and D.2 respectively.

D.1. Parameters for β-softmax-compare

Recall β-softmax-compare (Local Protocol 2) is the instatiation of the comparison dynamics where h is the following exponential function:

$$h_\beta(g) = \exp(\beta \cdot g) \quad \text{for all } g \in \mathbb{R},$$

for some $\beta \in (0, 1]$. Then for each $j \in [m]$, we can define (for a given β):

$$F_j(\mathbf{q}, \mathbf{g}) \;:=\; \sum_{k \in [m]} q_k \cdot \left(\frac{e^{\beta g_j} - e^{\beta g_k}}{e^{\beta g_j} + e^{\beta g_k}} \right). \tag{12}$$

We will now develop the proof of the following lemma, which gives parameter values that are sufficient for satisfying Assumption 2.

Lemma D.1 *Let $\mathcal{F} = \{F_j\}_{j \in [m]}$ be the zero-sum family induced by the β-softmax-compare protocol. Then for a reward sequence $\{\mathbf{g}^t\}$ with each $\mathbf{g}^t \in [-\sigma, \sigma]^m$, the family $\mathcal{F}$ satisfies Assumption 2 with parameters $\alpha_1 = \alpha_2 = \frac{3}{2}\beta$, $\delta = 4\beta\sigma$, and $L = 2$, for any $0 < \beta \leq 1/(4\sigma)$.*

To start, we will first develop the proof the the following lemma, which gives the α_1, α_2, and δ estimates for β-softmax-compare. Note that throughout, the function F_j refers specfically to the one induced by the β-softmax-compare dynamics from expression (12).

Lemma D.2 *For any $\mathbf{q} \in \Delta_m$ and $\mathbf{g} \in [-\sigma, \sigma]^m$, where $\mathbb{E}[\mathbf{g}] = \boldsymbol{\mu}$ and $\sigma \in [1, 10]$, it holds for all $\beta \in (0, \frac{1}{4}]$ that*

$$\frac{\beta}{2} \cdot \left|\mu_j - \langle \mathbf{q}, \boldsymbol{\mu} \rangle - 4\beta\sigma\right| \;\leq\; \left|\mathbb{E}_{\mathbf{q}}\left[F_j(\mathbf{q}, \mathbf{g})\right]\right| \;\leq\; \frac{\beta}{2} \cdot \left|\mu_j - \langle \mathbf{q}, \boldsymbol{\mu} \rangle + 4\beta\sigma\right| .$$

Note that in the above, and as mentioned in the problem setting from Section 1, we assume $\sigma \in [1, 10]$ for simplicity (and given that the means $\boldsymbol{\mu}$ are all bounded in $[-1, 1]$). Our technique yields similar bounds for larger σ at the expense of worse constants, and we omit the details.

The key step in proving Lemma D.2 is to first derive almost sure bounds on each term in F_j. Specifically, we first prove the following proposition, which "linearizes" the exponential differences.

Proposition D.3 *For any $\beta \leq \frac{1}{4}$ and for all $g_j, g_k \in [-10, 10]$:*

$$\Big(\frac{1}{2} - 2\beta\Big)\beta \cdot |g_j - g_k| \;\leq\; \left|\frac{e^{\beta g_j} - e^{\beta g_k}}{e^{\beta g_j} + e^{\beta g_k}}\right| \;\leq\; \frac{1}{2}\beta \cdot |g_j - g_k| \,.$$

In the proof, we leverage concavity (and convexity) properties of the exponential differences to derive linear approximations that have the appropriate upper or lower bound property.

Proof We begin by proving the upper bound (right hand inequality). For readability, let $x = g_j$, $y = g_k$, and define

$$\Phi(x, y) \;:=\; \frac{e^{\beta x} - e^{\beta y}}{e^{\beta x} + e^{\beta y}}$$

Our goal is to show that $|\Phi(x, y)| \leq \frac{1}{2}\beta|x - y|$, and by symmetry, it suffices to prove $\Phi(x, y) \leq \frac{1}{2}\beta(x - y)$ for all $x \geq y$ when $x \in [-10, 10]$.

Fixing y, we reduce this task to a one-dimensional argument in x. Differentiating, we find

$$\frac{\partial \Phi}{\partial x} = \frac{2\beta e^{\beta(x+y)}}{(e^{\beta x} + e^{\beta y})^2} \quad \text{and} \quad \frac{\partial^2 \Phi}{\partial x^2} = \frac{-2\beta^2 e^{\beta(x+y)}(e^{\beta x} - e^{\beta y})}{(e^{\beta x} + e^{\beta y})^3} \,.$$

Now observe that for $x \geq y$, this second derivative is always non-positive, meaning that $\Phi(x, y)$ is concave as a function of x. Then by concavity, the tangent line (with respect to x) through $x = y$ is an upper bound on Φ for all $x \geq y$. To define this tangent line, we can evaluate $\partial\Phi/\partial x$ at $x = y$, which gives a slope of $\beta/2$. Passing the line through the point $(y, \Phi(y, y)) = (y, 0)$ gives an intercept at $-(\beta/2)y$. Together, this implies the line $\frac{\beta}{2}(x - y)$ is an upper bound on $\Phi(x, y)$ for all $x \geq y$. Since this bound holds uniformly for any fixed y, it holds in general for all pairs $x \geq y$.

We now prove the lower bound (left hand inequality) from the lemma statement. For this, we start with the case when $\Phi(x, y) \geq 0$ (meaning $x \geq y$) and again fix some $y \in [-10, 10]$. Then recall from the proof of the right hand inequality above that $\Phi(x, y)$ is concave with respect to x in this domain. Thus when $x \geq y$, it follows by concavity that any secant line (with respect to x) passing through $(y, 0)$ and $(z, \Phi(z, y))$ (for $z \geq y$) is a lower bound on Φ.

Now consider the line $f(x, y) = ((1/2) - 2\beta) \cdot \beta(x - y)$, which is one such secant through $(y, 0)$. To show that $f(x, y)$ is a lower bound on $\Phi(x, y)$ when $-10 \leq y \leq x \leq 10$, it suffices to show that $\Phi(10, y) \geq f(10, y)$ for any $\beta \geq 0$, as this would imply (by concavity) that the intersection of $f(x, y)$ and $\Phi(x, y)$ occurs at some $x \geq 10$. For this, we can take the difference

$$\Phi(10, y) - f(10, y) \;=\; \frac{e^{\beta \cdot 10} - e^{\beta y}}{e^{\beta \cdot 10} + e^{\beta y}} - \big(\tfrac{1}{2} - 2\beta\big) \cdot \beta(10 - y)$$

and differentiate with respect to β to find that for any $-10 \leq y \leq x$, this difference is increasing for all $0 \leq \beta \leq 1/4$. Moreover, we observe that for all y, setting $\beta = 0$ yields $\Phi(10, y) - f(10, y) = 0$.

Together, this implies that the difference is non-negative for all $-10 \le y \le x \le 10$ and $0 \le \beta \le 1/4$, and the bound holds uniformly over all such x and y.

Now in the case that $\Phi(x,y)$ is negative (meaning $x < y$), our goal is to show $\Phi(x,y) \le ((1/2) - 2\beta) \cdot \beta(y-x)$. This can be accomplished by an identical argument as in the non-negative case, but instead leveraging the convexity (rather than concavity) of $\Phi(x,y)$ with respect to x in the domain $x < y$. We thus omit these repeated steps. Combining both cases, we conclude that $|\Phi(x,y)| \ge \left(\frac{1}{2} - 2\beta\right) \cdot \beta \cdot |x-y|$ for all $x, y \in [-10, 10]$. ■

Using the bounds in Proposition D.3, we can now prove Lemma D.2.

Proof *(of Lemma D.2)* We begin with the upper bound (right-hand side inequality). For this, fix $\mathbf{q} \in \Delta_m$, $\mathbf{g} \in [-\sigma, \sigma]^m$, $j \in [m]$, and $\beta \le 1/4$, and again define

$$\Phi(x,y) \;:=\; \frac{e^{\beta x} - e^{\beta y}}{e^{\beta x} + e^{\beta y}} \,.$$

To start, assume $F_j(\mathbf{q}, \mathbf{g}) \ge 0$ and define the sets

$$\begin{aligned} C^+ &= \{k \in [m] \;:\; \Phi(g_j, g_k) \ge 0\} \\ C^- &= \{k \in [m] \;:\; \Phi(g_j, g_k) < 0\} \,. \end{aligned}$$

Then we can write

$$F_j(\mathbf{q}, \mathbf{g}) \;=\; \sum_{k \in [m]} q_k \cdot \Phi(g_j, g_k) \;=\; \sum_{k \in C^+} q_k \cdot \Phi(g_j, g_k) - \sum_{k \in C^-} q_k \cdot \Phi(g_k, g_j) \,.$$

Now recall by definition that for any $x, y \in \mathbb{R}$, $\Phi \ge 0$ iff $x \ge y$. Then using the bounds on Φ from Proposition D.3 (which hold for $|g_j|, |g_k| \le \sigma \le 10$), observe that

$$\begin{aligned} \Phi(g_j, g_k) \;&\le\; \tfrac{\beta}{2}(g_j - g_k) \quad \text{for } k \in C^+ \\ \text{and} \quad \Phi(g_k, g_j) \;&\ge\; \tfrac{\beta}{2}(1 - 4\beta)(g_k - g_j) \quad \text{for } k \in C^- \,. \end{aligned}$$

It follows that we can bound F_j and rearrange to find

$$\begin{aligned} F_j(\mathbf{q}, \mathbf{g}) \;&\le\; \sum_{k \in C^+} q_k \cdot \tfrac{\beta}{2}(g_j - g_k) - \sum_{k \in C^-} q_k \cdot \tfrac{\beta}{2}(1 - 4\beta)(g_k - g_j) \\ &=\; g_j \Big(\sum_{k \in C^+} q_k \tfrac{\beta}{2} + \sum_{k \in C^-} q_k \tfrac{\beta}{2}(1 - 4\beta) \Big) - \Big(\sum_{k \in C^+} q_k g_k \cdot \tfrac{\beta}{2} + \sum_{k \in C^-} q_k g_k \tfrac{\beta}{2}(1 - 4\beta) \Big) \\ &\le\; \tfrac{\beta}{2} \cdot g_j - \tfrac{\beta}{2}(1 - 4\beta) \cdot \langle \mathbf{q}, \mathbf{g} \rangle \,, \end{aligned}$$

where in the final inequality we use the fact that $1 - 4\beta \le 1$. Then simplifying further and using Hölder's inequality gives

$$F_j(\mathbf{q}, \mathbf{g}) \;\le\; \frac{\beta}{2} \cdot \Big(g_j - \langle \mathbf{q}, \mathbf{g} \rangle + 4\beta \cdot \|\mathbf{g}\|_\infty \Big) \,,$$

which holds (almost surely) for $\mathbf{g} \in [-\sigma, \sigma]^m$ for $\sigma \in [1, 10]$. Then taking expectation conditioned on $\mathbf{q}$, we conclude

$$\mathbb{E}_{\mathbf{q}}\big[F_j(\mathbf{q}, \mathbf{g})\big] \;\le\; \frac{\beta}{2} \cdot \Big(\mu_j - \langle \mathbf{q}, \boldsymbol{\mu} \rangle + 4\beta\sigma \Big) \,.$$

In the case where $F_j(\mathbf{q}, \mathbf{g})$ is negative, we can apply the same argument to $-F_j(\mathbf{q}, \mathbf{g})$, which then proves the right-hand (upper bound) inequality of the lemma statement.

To prove the left-hand (lower bound) inequality of the lemma, a similar strategy as above will find (almost surely) that

$$\begin{aligned} F_j(\mathbf{q}, \mathbf{g}) \;&\geq\; \frac{\beta}{2} \cdot \big(g_j - \langle \mathbf{q}, \mathbf{g} \rangle - 4\beta \cdot |g_j|\big) \quad \text{when } F_j(\mathbf{q}, \mathbf{g}) \geq 0 \\ \text{and} \quad F_j(\mathbf{q}, \mathbf{g}) \;&\leq\; -\frac{\beta}{2} \cdot \big(g_j - \langle \mathbf{q}, \mathbf{g} \rangle - 4\beta \cdot |g_j|\big) \quad \text{when } F_j(\mathbf{q}, \mathbf{g}) < 0 \,, \end{aligned}$$

for all $j \in [m]$ under the assumptions on $\mathbf{g}$ and β from the lemma statement. Then taking expectation conditioned on $\mathbf{q}$ and noting that all $|g_j| \leq \sigma$ yields the desired left-hand bound. ■

With Proposition D.2 in hand, we can now prove Lemma D.1:

Proof [Proof (of Lemma D.1)] Let $\mathcal{F} = \{F_j\}_{j \in [m]}$ be the family induced by the β-softmax-compare dynamics, where each F_j is as defined in expression (12). Then using Lemma D.2, we can factor out a 3 to find that condition (i) of Assumption 2 is satisfied with $\alpha_1 = \alpha_2 = (3/2)\beta$, and $\delta = 4\sigma\beta$. We assume that $\beta \leq 1/6$ and $\beta \leq 1/(4\sigma)$, which ensures that both $\alpha_1 = \alpha_2 \leq 1/4$, and that $\delta \leq 1$.

For part (ii) of the assumption, observe that the range of each exponential difference, as defined in expression (12), is bounded in $[-2, 2]$, and thus $|F_j(\mathbf{q}, \mathbf{g}) - F_j(\mathbf{p}, \mathbf{g})| \leq 2\|\mathbf{q} - \mathbf{p}\|_1$ for any $\mathbf{q}, \mathbf{p} \in \Delta_m$ and each $j \in [m]$. Thus setting $L = 2$ allows the assumption to be satisfied. ■

D.2. Parameters for β-sigmoid-adopt

Recall that β-sigmoid-adopt (Local Protocol 1) is the instantiation of an adoption dynamics with the following sigmoid adoption function $f_\beta(g) = \frac{1}{1+\exp(-\beta \cdot g)}$, where $\beta \in (0, 1]$. Then for each $j \in [m]$, we have from Proposition A.1 that

$$F_j(\mathbf{q}, \mathbf{g}) \;:=\; \sum_{k \in [m]} q_k \cdot \left(\frac{1}{1 + e^{-\beta \cdot g_j}} - \frac{1}{1 + e^{-\beta \cdot g_k}} \right) . \tag{13}$$

Then the following lemma establishes the parameter values under which β-sigmoid-adopt satisfies Assumption 2.

Lemma D.4 *Let $\mathcal{F} = \{F_j\}_{j \in [m]}$ be the zero-sum family induced by the β-sigmoid-adopt protocol. Then for a reward sequence $\{\mathbf{g}^t\}$ with each $\mathbf{g}^t \in [-\sigma, \sigma]^m$, the family $\mathcal{F}$ satisfies Assumption 2 with parameters $\alpha_1 = \alpha_2 := \frac{3}{4}\beta$, $\delta := 4\beta\sigma$, and $L := 2$, for any $\beta \leq \min\{\frac{1}{4\sigma}, \frac{1}{3}\}$.*

We start by deriving the parameters α_1, α_2 and δ. For this, in the following proposition, we establish (almost surely) two-sided bounds on the sigmoid function that are linear in g_j (analogous to the linear bounds on the softmax differences from Proposition D.3).

Proposition D.5 *For any $\beta \in (0, \frac{1}{4}]$:*

$$\begin{aligned} \tfrac{1}{2} + \left(\tfrac{\beta}{4} - \beta^2\right) \cdot x \;&\leq\; \frac{1}{1 + e^{-\beta \cdot x}} \;\leq\; \tfrac{1}{2} + \tfrac{\beta}{4} \cdot x \quad \textit{for } x \in [0, 10] \\ \tfrac{1}{2} + \left(\tfrac{\beta}{4} - \beta^2\right) \cdot x \;&\geq\; \frac{1}{1 + e^{-\beta \cdot x}} \;\geq\; \tfrac{1}{2} + \tfrac{\beta}{4} \cdot x \quad \textit{for } x \in [-10, 0) \,. \end{aligned}$$

Proof The proof technique is similar to that of Proposition D.3. We will verify the bounds for $x \in [0, 10]$ as the bounds for $x \in [-10, 0)$ will follow by symmetry.

Thus consider $x \in [0, 10]$. For the upper bound, it suffices to show that

$$\tfrac{1}{2} + \tfrac{\beta}{4}x - \frac{1}{1 + e^{-\beta x}} \geq 0$$

for all $\beta \in (0, \frac{1}{4}]$. Observe that the difference is exactly 0 when $\beta = 0$, and by differentiating with respect to β, one can verify that the difference is increasing for $\beta \leq \frac{1}{4}$ when $x \geq 0$. This establishes the upper bound.

For the lower bound, we apply the same reasoning to the difference

$$\frac{1}{1 + e^{-\beta x}} - \tfrac{1}{2} - \tfrac{\beta}{4}x + \beta^2 x \;\geq\; 0$$

for all $\beta \in (0, \frac{1}{4}]$, when $0 \leq x \leq 10$. Again observe that the difference is 0 when $\beta = 0$, and we can differentiate with respect to β to find that the difference is increasing for all $\beta \leq \frac{1}{4}$ when $x \leq 10$. ■

Using Proposition D.5, we can then state the following inequalities with respect to $|\mathbb{E}_{\mathbf{q}}[F_j(\mathbf{q}, \mathbf{g})]|$:

Lemma D.6 *Consider the family $\mathcal{F} = \{F_j\}_{j \in [m]}$, where each F_j is defined as in expression (13) with parameter β. Then for any $\mathbf{q} \in \Delta_m$ and $\mathbf{g} \in [-\sigma, \sigma]^m$, where $\mathbb{E}[\mathbf{g}] = \boldsymbol{\mu}$ and $\sigma \in [1, 10]$, it holds for all $\beta \in (0, \frac{1}{4}]$ that*

$$\frac{\beta}{4} \cdot \left|\mu_j - \langle \mathbf{q}, \boldsymbol{\mu} \rangle - 4\beta\sigma\right| \;\leq\; \left|\mathbb{E}_{\mathbf{q}}\left[F_j(\mathbf{q}, \mathbf{g})\right]\right| \;\leq\; \frac{\beta}{4} \cdot \left|\mu_j - \langle \mathbf{q}, \boldsymbol{\mu} \rangle + 4\beta\sigma\right| .$$

The key step in the proof is to observe that for fixed β any $-\sigma \leq g_j \leq g_k \leq \sigma$:

$$\begin{aligned} \frac{1}{1 + e^{-\beta g_j}} - \frac{1}{1 + e^{-\beta g_k}} \;&\leq\; \frac{\beta}{4}(g_j - g_k) + \beta^2\sigma \\ \text{and} \quad \frac{1}{1 + e^{-\beta g_j}} - \frac{1}{1 + e^{-\beta g_k}} \;&\geq\; \frac{\beta}{4}(g_j - g_k) - \beta^2\sigma \,, \end{aligned}$$

which follows (almost surely) from Proposition D.5. From here, we use an identical strategy as in Lemma D.2 to account for the positive and negative terms in the summation in F_j, and then take conditional expectations to derive the final bounds. As the remainder of the proof follows identically to that of Lemma D.2, we omit these details.

Proof of Lemma D.4 With Lemma D.6 in hand, the proof of Lemma D.4 follows identically to that of Lemma D.1: First, we use the inequalities of Lemma D.6 and factor out a 3 to establish the $\alpha_1 = \alpha_2$ and δ parameters. Then, we again use the observation that the range of each sigmoid difference is bounded in $[-2, 2]$, and thus setting $L = 2$ suffices to satify condition (ii) of the assumption.

Appendix E. Details on Coupling Error Anlaysis

In this appendix, we develop the proof of Lemma B.3 (restated below) which bounds the error on the coupling from Definition 2.2:

Lemma B.3 *Consider the sequences $\{\mathbf{p}^t\}$, $\{\widehat{\mathbf{p}}^t\}$, and $\{\mathbf{q}^t\}$ from Definition 2.2 with a reward sequence $\{\mathbf{g}^t\}$ and using a family $\mathcal{F}$ that satisfies Assumption 2 with parameter L. Let $\kappa := (3 + L)$, and assume $n \geq 3c \log n$ for some $c \geq 1$. Then for any $T \geq 1$:*

$$\sum_{t\in[T]} \mathbb{E}\|\mathbf{q}^{t+1} - \mathbf{p}^{t+1}\|_1 \;\leq\; \widetilde{O}\Big(\frac{m \cdot \kappa^T}{\sqrt{n}} + \frac{m \cdot T}{n^c}\Big)\,.$$

We start by sketching an overview of the argument. First, recall by the law of iterated expectation that for each $t \in [T]$:

$$\mathbb{E}\|\mathbf{q}^{t+1} - \mathbf{p}^{t+1}\|_1 \;=\; \mathbb{E}\big[\,\mathbb{E}_t\|\mathbf{q}^{t+1} - \mathbf{p}^{t+1}\|_1\,\big]\,.$$

Then by the triangle inequality and linearity of expectation, it follows that

$$\begin{aligned}\sum_{t\in[T]} \mathbb{E}\|\mathbf{q}^{t+1} - \mathbf{p}^{t+1}\|_1 \;&\leq\; \mathbb{E}\Big[\sum_{t\in[T]} \mathbb{E}_t\|\mathbf{q}^{t+1} - \widehat{\mathbf{p}}^{t+1}\|_1 + \mathbb{E}_t\|\widehat{\mathbf{p}}^{t+1} - \mathbf{p}^{t+1}\|_1\Big] \\ &=\; \mathbb{E}\Big[\sum_{t\in[T]} \mathbb{E}_t\|\mathbf{q}^{t+1} - \widehat{\mathbf{p}}^{t+1}\|_1 + \|\widehat{\mathbf{p}}^{t+1} - \mathbf{p}^{t+1}\|_1\Big]\,. \qquad (14)\end{aligned}$$

Here, the final equality is due to the fact that both $\widehat{\mathbf{p}}^{t+1}$ and $\mathbf{p}^{t+1}$ are functions of $\{\mathbf{p}^t\}$ and $\{\mathbf{g}^t\}$, which means $\mathbb{E}_t\|\widehat{\mathbf{p}}^{t+1} - \mathbf{p}^{t+1}\|_1 = \|\widehat{\mathbf{p}}^{t+1} - \mathbf{p}^{t+1}\|_1$ for each t.

Thus in expression (14), we have decomposed the error (in conditional expectation) at each round t as the sum of the distances between $\mathbf{q}^t$ and $\widehat{\mathbf{p}}^t$ and $\widehat{\mathbf{p}}^t$ and $\mathbf{p}^t$. For the former, recall that $\mathbf{q}^t$ and $\widehat{\mathbf{p}}^t$ are related under the randomness of $\mathbf{g}^t$ and the same zero-sum family $\mathcal{F} = \{F_j\}_{j\in[m]}$. Thus if $\mathbf{q}^{t-1}$ and $\mathbf{p}^{t-1}$ are close, we intuitively expect $\mathbf{q}^t$ and $\widehat{\mathbf{p}}^t$ to also be close. For the latter, oberseve that this distance is simply the deviation of $\mathbf{p}^t$ from its (conditional) mean $\widehat{\mathbf{p}}^t$, which can be controlled using a Chernoff bound. We make this intuition precise via the following two propositions:

Proposition E.1 *For every $t \geq 1$:* $\mathbb{E}_t\big\|\widehat{\mathbf{p}}^{t+1} - \mathbf{q}^{t+1}\big\|_1 \;\leq\; (2 + L) \cdot \mathbb{E}_{t-1}\big\|\mathbf{p}^t - \mathbf{q}^t\big\|_1$.

Proposition E.2 *For any $c \geq 1$ and $n \geq 3c \log n$, it holds for every $t \in [T]$ simultaneously that*

$$\big\|\mathbf{p}^t - \widehat{\mathbf{p}}^t\big\|_1 \;\leq\; m \cdot \sqrt{\frac{3c \log n}{n}}$$

with probability at least $1 - \frac{2mT}{n^c}$.

Granting both propositions true for now, we can then prove the main lemma:

Proof *(of Lemma B.3)* Fix $c \geq 1$ and assume $n \geq 3c \log n$. By substituting the bound of Proposition E.2 into expression (14), we find that

$$\begin{aligned}\mathbb{E}_t\|\mathbf{q}^{t+1} - \mathbf{p}^{t+1}\|_1 \;&\leq\; \mathbb{E}_t\big\|\mathbf{q}^{t+1} - \widehat{\mathbf{p}}^{t+1}\big\|_1 + \big\|\widehat{\mathbf{p}}^{t+1} - \mathbf{p}^{t+1}\big\|_1 \\ &\leq\; \mathbb{E}_t\big\|\mathbf{q}^{t+1} - \widehat{\mathbf{p}}^{t+1}\big\|_1 + \frac{m\sqrt{3c \log n}}{\sqrt{n}} \qquad (15)\end{aligned}$$

for all $t \in [T]$ simultaneously with probability at least $1 - \frac{2(T+1)}{n^c}$. Then substituting the bound of Proposition E.1 into expression (15), for each t we find

$$\sum_{t\in[T]} \mathbb{E}_t\|\mathbf{q}^{t+1} - \mathbf{p}^{t+1}\|_1 \;\le\; (2+L)\cdot \mathbb{E}_{t-1}\|\mathbf{q}^t - \mathbf{p}^t\|_1 + \frac{m\sqrt{3c\log n}}{\sqrt{n}}$$

simultaneously with probability at least $1 - \frac{2(T+1)}{n^c}$. Now recall by definition that $\mathbf{p}^0 = \mathbf{q}^0$, which implies $\mathbb{E}_0[\mathbf{q}^1] = \mathbb{E}_0[\widehat{\mathbf{p}}^1]$. Then unrolling the recurrence yields

$$\mathbb{E}_t\|\mathbf{q}^{t+1} - \mathbf{p}^{t+1}\|_1 \;\le\; (3+L)^t \cdot \frac{m\sqrt{3c\log n}}{\sqrt{n}}$$

for each $t \in [T]$, again with probability at least $1 - \frac{2(T+1)}{n^c}$. Reindexing and summing over all t yields with probability at least $1 - \frac{2T}{n^c}$:

$$\sum_{t\in[T]} \mathbb{E}_{t-1}\|\mathbf{q}^t - \mathbf{p}^t\|_1 \;\le\; \sum_{t\in[T]} (3+L)^{t-1}\cdot \frac{m\sqrt{3c\log n}}{\sqrt{n}} \;\le\; (3+L)^T \cdot \frac{m\sqrt{3c\log n}}{\sqrt{n}}\,.$$

Finally, taking expectations, we conclude

$$\sum_{t\in[T]} \mathbb{E}\|\mathbf{q}^t - \mathbf{p}^t\|_1 \;\le\; (3+L)^T \cdot \frac{m\sqrt{3c\log n}}{\sqrt{n}} + \frac{2mT}{n^c}\,.$$

Hiding the leading constants and logarithmic dependence on n in the $\widetilde{O}(\cdot)$ expression completes the proof of the lemma. ■

It now remains to prove Propositions E.1 and E.2, which we do in the following subsections.

E.1. Proof of Proposition E.1

For convenience, we restate the proposition:

Proposition E.1 *For every $t \ge 1$:* $\mathbb{E}_t\big\|\widehat{\mathbf{p}}^{t+1} - \mathbf{q}^{t+1}\big\|_1 \;\le\; (2+L)\cdot \mathbb{E}_{t-1}\big\|\mathbf{p}^t - \mathbf{q}^t\big\|_1$.

Proof Recall by definition that

$$\begin{aligned} q_j^{t+1} &= q_j^t \cdot \big(1 + F_j(\mathbf{q}^t, \mathbf{g}^t)\big) \\ \text{and } \widehat{p}_j^{t+1} &= p_j^t \cdot \big(1 + F_j(\mathbf{p}^t, \mathbf{g}^t)\big) \end{aligned}$$

for all $j \in [m]$. For readability we will write $\widehat{\mathbf{p}}', \mathbf{p}', \mathbf{q}'$ for $\widehat{\mathbf{p}}^{t+1}$, $\mathbf{p}^{t+1}$, $\mathbf{q}^{t+1}$, and $\mathbf{p}$, $\mathbf{q}$, $\mathbf{g}$ for $\mathbf{p}^t$, $\mathbf{q}^t$, $\mathbf{g}^t$, respectively. It follows that

$$\begin{aligned} \mathbb{E}_t\big\|\widehat{\mathbf{p}}' - \mathbf{q}'\big\|_1 &= \sum_{j\in[m]} \mathbb{E}_t\big|p_j - q_j + p_j\cdot F_j(\mathbf{p},\mathbf{g}) - q_j\cdot F_j(\mathbf{q},\mathbf{g})\big| \\ &\le \sum_{j\in[m]} \mathbb{E}_t|p_j - q_j| + \mathbb{E}_t|(p_j - q_j)\cdot F_j(\mathbf{p},\mathbf{g})| + \mathbb{E}_t|q_j\cdot(F_j(\mathbf{p},\mathbf{g}) - F_j(\mathbf{q},\mathbf{g}))| \\ &\le \mathbb{E}_t\|\mathbf{p}-\mathbf{q}\|_1 + \sum_{j\in[m]} \mathbb{E}_t|p_j - q_j| + q_j\big(L\cdot\mathbb{E}_t\|\mathbf{p}-\mathbf{q}\|_1\big) \\ &= (2+L)\cdot\mathbb{E}_t\|\mathbf{p}-\mathbf{q}\|_1\,. \end{aligned}$$

Here, the first line follows from two applications of the triangle inequality, and the second line comes from applying the boundedness and L-Lipschitz property of each F_j from Definition 2.1 and part (ii) of Assumption 2.

Finally, because $\mathbf{p}^t = \mathbf{p}$ and $\mathbf{q}^t = \mathbf{q}$ are functions only of $\{\mathbf{p}^{t-1}\}$ and $\{\mathbf{g}^{t-1}\}$, it follows that $\mathbb{E}_t\|\mathbf{p} - \mathbf{q}\|_1 = \mathbb{E}_{t-1}\|\mathbf{p} - \mathbf{q}\|$. Thus we conclude that $\mathbb{E}_t\|\widehat{\mathbf{p}}' - \mathbf{q}'\|_1 \le (2 + L) \cdot \mathbb{E}_{t-1}\|\mathbf{p} - \mathbf{q}\|_1$. ■

E.2. Proof of Proposition E.2

For convenience, we restate the proposition:

Proposition E.2 *For any $c \ge 1$ and $n \ge 3c \log n$, it holds for every $t \in [T]$ simultaneously that*

$$\left\|\mathbf{p}^t - \widehat{\mathbf{p}}^t\right\|_1 \;\le\; m \cdot \sqrt{\frac{3c \log n}{n}}$$

with probability at least $1 - \frac{2mT}{n^c}$.

Proof Using a standard multiplicative Chernoff bound [24, Corollary 4.6], we have for each $j \in [m]$ and $t \in [T]$ that

$$\mathbb{P}_{t-1}\Big(\Big| \, p_j^t - \mathbb{E}_{t-1}[p_j^t] \, \Big| \ge \mathbb{E}_{t-1}[p_j^t] \cdot \delta \Big) \;\le\; 2 \cdot \exp\Big(-\frac{n}{3} \cdot \mathbb{E}_{t-1}[p_j^t] \cdot \delta^2 \Big) \,,$$

for any $0 < \delta \le 1$. Fix $c \ge 1$, and consider the case when $\sqrt{\frac{3c \log n}{n}} \le \mathbb{E}_{t-1}[p_j^t] \le 1$. Then setting $\delta = \frac{1}{\mathbb{E}_{t-1}[p_j^t]} \cdot \sqrt{\frac{3c \log n}{n}} \le 1$ implies

$$\mathbb{P}_{t-1}\Big(\Big| \, p_j^t - \mathbb{E}_{t-1}[p_j^t] \, \Big| \ge \sqrt{\tfrac{3c \log n}{n}} \Big) \;\le\; 2 \cdot \exp\Big(\tfrac{-c \log n}{\mathbb{E}_{t-1}[p_j^t]} \Big) \;\le\; \frac{2}{n^c} \,.$$

On the other hand, when $0 \le \mathbb{E}_{t-1}[p_j^t] < \sqrt{\frac{3c \log n}{n}}$, setting $\delta = 1$ implies

$$\mathbb{P}_{t-1}\Big(\Big| \, p_j^t - \mathbb{E}_{t-1}[p_j^t] \, \Big| \ge \sqrt{\tfrac{3c \log n}{n}} \Big) \;\le\; 2 \cdot \exp\Big(-\tfrac{1}{3} \cdot \sqrt{3cn \log n} \Big) \;\le\; \frac{2}{n^c} \,,$$

where the final inequality holds for all $n \ge 3c \log n$.

Summing over all m coordinates, T rounds, and taking a union bound concludes the proof. ■

Appendix F. Details on Instantiated Regret Bounds

In this section, we provide details on the instantiated regret bounds from Section 2.3. Specifically, we develop the proof of Theorem 2.3 (restated below), which gives a regret bound for the β-softmax-compare and β-sigmoid-adopt dynamics:

Theorem 2.3 *Consider the sequence $\{\mathbf{p}^t\}$ induced by running the β-softmax-compare or β-sigmoid-adopt protocol on an (adversarial) reward sequence $\{\mathbf{g}^t\}$ intialized from $\mathbf{p}^0 = \mathbf{1}/m$. Then for any $c \geq 1$ and $n \geq 3c \log n$, and assuming that $T \leq (\frac{1}{2} - \epsilon) \log_5 n = O\big(\log\big(\frac{n}{m^2 \log n}\big)\big)$ for some $\epsilon \in (0, \frac{1}{2})$, setting $\beta := \sqrt{(\log m)/T}$ yields average regret of:*

$$\frac{1}{T} \cdot R(T) \;\leq\; O\left(\sqrt{\frac{\log m}{T}}\right) + \widetilde{O}\left(\frac{\sigma m}{n^{\epsilon}} + \frac{\sigma m}{n^{c}}\right).$$

To develop the proof of the theorem, we start by establishing a general T-step regret that stems from the framework introduced in Section 2.2. We then derive a bound on the worst-case mass decay of any arm, which establishes a bound on the time horizon T for which sub-linear regret bounds in this adversarial setting can be obtained.

F.1. General T-step Regret Bound

First, recall from Section 2.2 that for a sequence $\{\mathbf{p}^t\}$ as defined in the coupling of Definition 2.2, we can use the zero-sum MWU regret bound of Theorem B.2 and the coupling error bound of Lemma B.3 to derive an overall T-round regret bound. We state this bound more formally, which leverages the fact (expression (6) and Proposition B.1) that $R(T) \leq \widehat{R}(T)$:

Proposition F.1 *Consider the sequence $\{\mathbf{p}^t\}$ as defined in the coupling of Definition 2.2 and using a family $\mathcal{F}$ that satisfies Assumption 2 with parameters α_1, α_2, and δ. Moreover, assume that $\alpha = \alpha_2 = \alpha$, and that $\delta = O(\alpha)$, and that the reward sequence $\{\mathbf{g}^t\}$ is such that each $\mathbf{g}^t \in [-\sigma, \sigma]^m$. Then intialized from $\mathbf{p}^0 = \mathbf{1}/m$, for any $T \geq 1$:*

$$R(T) \;\leq\; O\left(\frac{\log m}{\alpha} + \alpha T\right) + \widetilde{O}\left(\frac{\sigma m \kappa^T}{\sqrt{n}} + \frac{\sigma m T}{n^{c}}\right).$$

The proof of the proposition follows directly from applying the zero-sum MWU regret bound of Theorem B.2 and the coupling error bound of Lemma B.3 to expression (6). We also make the following remarks:

Remark F.2 *In the statement of the proposition, the dependence on σ comes from the regret decomposition in expression (6), and that the $\widetilde{O}(\cdot)$ notation hides only a $\sqrt{\log n}$ in the first term, and a $\sqrt{\log m}$ dependence in the second term, both of which we assume are dominated by their respective denominators. Additionally, while we assume the (α, δ) parameters of $\mathcal{F}$ have certain "nice" properties (which are satisfied by the corresponding families induced by our local dynamics), one can derive similar T-round regret bounds using this framework for any zero-sum family $\mathcal{F}$, but with different (larger) dependencies on α_1, α_2, and δ. Thus given some family $\mathcal{F}$, if one can establish tighter two-sided bounds on the magnitude of each $\mathbb{E}_{\mathbf{q}}[F_j(\mathbf{q}, \mathbf{g})]$ with respect to $\mu_j - \langle \mathbf{q}, \boldsymbol{\mu} \rangle$ (i.e., showing $\alpha_2 - \alpha_1 = 0$ and that δ is small), then tighter regret bounds can be obtained.*

Remark F.3 (Applying the Bound to j'th Arm Regret) *We remark that the T-round regret bound in Proposition F.1 (as well as the zero-sum MWU regret bound of Theorem B.2) can also be stated more generally with respect to any arm j that initially satisfies the requisite mass lower bound constraint (i.e., $p_j^t \geq 1/\rho$). To see this, observe that the only dependence on j in the decomposition of $\widehat{R}(T)$ (i.e., from Proposition B.1) comes from the zero-sum MWU bound on $\{\mathbf{q}^t\}$, which requires a*

lower bound ρ on the initial mass q_j^0. Thus if ρ is a (probabilistic) uniform lower bound on the mass of every coordinate j at round 0, then it follows that the bound in Proposition F.1 also applies more generally to the "j'th-arm regret" of $\sum_{t\in[T]} \mu_j^t - \sum_{t\in[T]} \mathbb{E}[\langle \mathbf{p}^t, \mathbf{g}^t\rangle]$.

F.2. Worst Case Mass Decay for any Arm

As mentioned in Section 2.3, in this adversarial reward setting, we require establishing how quickly the adoption mass of any arm can decay. This translates into a constraint on how large (with high probability) the number of rounds T can grow while still ensuring that every arm has at least $1/n$ adoption mass at every round. Thus we start by bounding this worst-case mass decay in the adversarial reward setting.

For this, note that when we have no additional assumptions about how the reward sequence is generated, we can only make a very pessimistic estimate about the size of the adoption mass of any arm j. In particular, even for the arm c maximizing $\max_{j\in[m]} \sum_{t\in[T]} \mu_j^t$, the weight p_c^{t+1} can be maximally decreasing with respect to p_c^t at any given round. Thus in the following lemma, we quantify this worst-case decay at any coordinate after t iterations.

Proposition F.4 *Consider the trajectory $\{\mathbf{p}^t\}$ from Definition 2.2 with an arbitrary reward sequence $\{\mathbf{g}^t\}$ running with a family $\mathcal{F} = \{F_j\}_{j\in[m]}$ that satisfies Assumption 2 with parameters α_1, α_2, and δ. Then for any $t \geq 1$ and $c \geq 1$ it holds for any $j \in [m]$ that*

$$p_j^{t+1} \;\geq\; p_j^0 \cdot \left(\frac{3}{4}\right)^t - \frac{4}{3}\sqrt{\frac{3c\log n}{n}}t$$

with probability at least $1 - \frac{2t}{n^c}$ when $n \geq 3c\log n$.

Proof Fix $j \in [m]$. To start, we use the update rule of $\mathbb{E}_t[p_j^{t+1}]$ and take expectation with respect to $\mathbf{g}^t$ to write

$$\mathbb{E}_{\mathbf{p}^t}[p_j^{t+1}] \;=\; p_j^t \cdot \Big(1 + \mathbb{E}_{\mathbf{p}}^t\big[F_j(\mathbf{p}^t, \mathbf{g}^t)\big]\Big) \;\geq\; p_j^t \cdot \Big(1 - \tfrac{\alpha_2}{3}\big|\mu_j^t - \langle \mathbf{p}, \boldsymbol{\mu}\rangle + \delta\big|\Big)\,,$$

where the inequality follows from the (worst-case) assumption that $\mathbb{E}_{\mathbf{p}}^t[F_j(\mathbf{p}^t, \mathbf{g}^t)] < 0$ and applying the bound from Assumption 2. Now under the assumptions that $|\mu_j^t| \leq 1$ and $\delta \leq 1$, it follows that $|\mu_j^t - \langle \mathbf{p}^t, \mathbf{g}^t\rangle + \delta| \leq 3$ for any $\mathbf{p}^t$ and $\mathbf{g}^t$. Together with the fact that $\alpha_2 \leq \frac{1}{4}$ by assumption, we can write

$$\mathbb{E}_{\mathbf{p}^t}[p_j^{t+1}] \;\geq\; p_j^t \cdot \Big(1 - \alpha_2\Big) \;\geq\; \frac{3}{4} \cdot p_j^t\,,$$

and by a Chernoff bound argument (i.e., applying the argument of Proposition E.2 at a single coordinate), we find that

$$p_j^{t+1} \;\geq\; \frac{3}{4} \cdot p_j^t - \sqrt{\frac{3c\log n}{n}}$$

with probability at least $1 - \frac{2}{n^c}$ for any $c \geq 1$ when $n \geq 3c\log n$. Then starting from the vector $\mathbf{p}^0$ at round $t = 0$, we can repeat this argument t times to find

$$p_j^{t+1} \;\geq\; p_j^0 \cdot \left(\frac{3}{4}\right)^t - \sqrt{\frac{3c\log n}{n}} \cdot \left(\sum_{i\in[t]} \left(\frac{3}{4}\right)^{i-1}\right) \;\geq\; p_j^0 \cdot \left(\frac{3}{4}\right)^t - \frac{4}{3}\sqrt{\frac{3c\log n}{n}}\,,$$

with probability at least $1 - \frac{2t}{n^c}$, where the second inequality follows by underestimating the negative term by an infinite geometric series. ■

Using this worst-case decay, we can derive a (pessismistic) upper bound on the number of rounds T for which, with high probability, $p_j^T \geq \frac{1}{n}$ (i.e., at least one node adopts every arm):

Proposition F.5 *Consider the trajectory $\{\mathbf{p}^t\}$ from Definition 2.2 with an arbitrary reward sequence $\{\mathbf{g}^t\}$ running with a family $\mathcal{F} = \{F_j\}_{j\in[m]}$ that satisfies Assumption 2 with parameters α_1, α_2, and δ. Assume that $\mathbf{p}^0 = \frac{1}{m}\mathbf{1}$ with probability 1. Then for any $T \leq \log\left(\frac{3n}{64cm^2 \log n}\right)$, it holds for every $j \in [m]$ and $t \in [T]$ that $p_j^{t+1} \geq \frac{1}{n}$ with probability at least $1 - \frac{2tm}{n^c}$, for any $c \geq 1$ and $n \geq 3c \log n$.*

Proof Using Proposition F.4, we have for any $j \in [m]$ that

$$p_j^{t+1} \;\geq\; \frac{1}{m} \cdot \left(\frac{3}{4}\right)^t - \frac{4}{3}\sqrt{\frac{3c \log n}{n}}\,. \tag{16}$$

Now suppose that we have t, n, and m satisfying

$$\frac{4}{3}\sqrt{\frac{3c \log n}{n}} \;\leq\; \frac{1}{m} \cdot \left(\frac{3}{4}\right)^t. \tag{17}$$

Then it follows from expression (16) that $p_j^{t+1} \geq \frac{1}{n}$ with probability at least $1 - \frac{2t}{n^c}$ as long as $t \leq \log\left(\frac{n}{2m}\right)/\log(4/3) =: T_a$. Now checking the constraint induced by (17), we find that

$$t \leq \log\left(\frac{3n}{64cm^2 \log n}\right) =: T_b$$

is sufficient to ensure this inequality holds. Thus observing that $T_b \leq T_a$ for all $n, m, c \geq 1$, it follows that constraining $t \leq T_b$ is sufficient to ensure that $p_j^{t+1} \geq \frac{1}{n}$ with probability at least $1 - (2t/n^c)$. Taking a union bound over all m coordinates yields the statement of the lemma. ■

F.3. Deriving the Final Regret Bound

Now using the time horizon constraint from Proposition F.5, we can prove Theorem 2.3. For this, we note that following analysis tradeoff: on the one hand, Proposition F.5 establishes a uniform lower bound of $1/n$ on every arm j so long as the constraint on T is satisfied. In particular, we specify that $T \leq (\frac{1}{2} - \epsilon)\log_\kappa n$, where $\epsilon \in (0, \frac{1}{2})$ is a tunable parameter satisfying $(\frac{1}{2} - \epsilon)\log_\kappa n = O(\log(n/(m^2 \log n)))$ (i.e., the constraint from Proposition F.5). Thus in the proof of Theorem 2.3, we simply apply the T-step regret bound of Proposition F.1 using this constraint on T, and we then tune the free parameter of our protocols accordingly:

Proof *(of Theorem 2.3)* We use the T round regret bound from Proposition F.1. First, recall by Lemmas D.1 and D.4 that the family $\mathcal{F}$ induced by these protocols each satisfy Assumption 2 with the following parameters:

- For β-softmax-compare, $\alpha_1 = \alpha_2 = \frac{3}{2}\beta$, $\delta = 4\beta\sigma$, and $L = 2$.

- For β-sigmoid-adopt, $\alpha_1 = \alpha_2 = \frac{3}{4}\beta$, $\delta = 4\beta\sigma$, and $L = 2$.

Thus the parameters associated with each protocol satisfy Assumption 2 with parameters $\alpha_1 = \alpha_2 = O(\beta)$, $\delta = O(\beta)$, and $L = 2$. Thus applying the bound from Lemma F.1 with $\alpha = O(\beta)$ and dividing by T shows that

$$\begin{aligned}\frac{1}{T}\cdot R(T) \;&\le\; O\left(\frac{\log m}{T\cdot\beta}+\beta\right)+\widetilde{O}\left(\frac{\sigma m\kappa^T}{T\sqrt{n}}+\frac{\sigma m}{n^c}\right)\\ &\le\; O\left(\frac{\log m}{T\cdot\beta}+\beta\right)+\widetilde{O}\left(\frac{\sigma m}{n^\epsilon}+\frac{\sigma m}{n^c}\right).\end{aligned}$$

Here, the final line comes from the assumption that $T \le ((1/2)-\epsilon)\log_\kappa n$ for some $\epsilon \in (0, \frac{1}{2})$ and for $\kappa = 3 + L = 5$, and thus $\kappa^T/(T\sqrt{n}) \le O(1/n^\epsilon)$. Finally setting $\beta := \sqrt{(\log m)/T}$ and simplifying the first term yields

$$\frac{1}{T}\cdot R(T) \;\le\; O\left(\sqrt{\frac{\log m}{T}}\right)+\widetilde{O}\left(\frac{\sigma m}{n^\epsilon}+\frac{\sigma m}{n^c}\right),$$

which concludes the proof. ■

Appendix G. Details on Convex Optimization Application

Here, we develop the proof of Theorem 2.4, which gives an error rate on the regret obtained using our comparison and adoption dynamics to approximately optimize a convex function $f : \Delta_m \to \mathbb{R}$ when the reward sequence $\{\mathbf{g}^t\}$ is generated using a stochastic gradient oracle as specified in Assumption 1. For convenience, we restate the theorem here:

Theorem 2.4 *Given a convex function $f : \Delta_m \to \mathbb{R}$, consider the sequence $\{\mathbf{p}^t\}$ induced by running the β-softmax-compare or β-sigmoid-adopt protocol on a reward sequence $\{\mathbf{g}^t\}$ generated as in Assumption 1 with gradient bound G. Then for any $c \ge 1$ and $n \ge 3c\log n$, assume that $T \le (\frac{1}{2}-\epsilon)\log_5 n = O\big(\log\big(\frac{n}{m^2\log n}\big)\big)$ for some $\epsilon \in (0, \frac{1}{2})$, and set $\beta := \sqrt{(\log m)/T}$. Let $\widetilde{\mathbf{p}} := \frac{1}{T}\sum_{t\in[T]}\mathbf{p}^t$ denote the average arm distribution over T rounds. Then:*

$$\mathbb{E}[f(\widetilde{\mathbf{p}})] - \min_{\mathbf{p}\in\Delta_m} f(\mathbf{p}) \;\le\; O\left(\sqrt{\frac{G^2\log m}{T}}\right)+\widetilde{O}\left(G\cdot\left(\frac{\sigma m}{n^\epsilon}+\frac{\sigma m}{n^c}\right)\right).$$

First, we note that this error rate is equivalent to our regret bound from the adversarial setting up to the factor G, which is a standard dependence. Note also that the optimization error is defined *implicitly*: the function f is being minimized with respect to the distribution $\mathbf{p}^t$ induced by the local dynamics. This is contrast to other settings of gossip-based, decentralized optimization (e.g., [16, 26, 32]), where each node $i \in [n]$ has first-order gradient access to an individual local function f_i, and the population seeks to perform empirical risk minimization over the n functions.

Now in order to prove the theorem, we first require relating the regret of the trajectory $\{\mathbf{p}^t\}$ to the expected primal gap $\mathbb{E}[f(\mathbf{p}^t) - f(\mathbf{p}^\star)]$ where $\mathbf{p}^\star \in \Delta_m$ is a function minimizer of f. For this, we give the following lemma, which follows similarly to that of Arora et al. [3, Theorem 3.11], but is adapted to handle stochastic rewards.

Lemma G.1 *Let $\{\mathbf{p}^t\}$ be a sequence of distributions, and let $\{\mathbf{g}^t\}$ be a sequence of rewards generated as in Assumption 1 with respect to a convex function f and gradient bound G. Then for any $T \geq 1$, letting $\widehat{\mathbf{p}} := \frac{1}{T}\sum_{t\in[T]} \mathbf{p}^t$ and $\mathbf{p}^\star := \arg\max_{p\in\Delta_m} f(\mathbf{p})$:*

$$\mathbb{E}[f(\widetilde{\mathbf{p}})] - f(\mathbf{p}^\star) \;\leq\; G \sum_{j\in[m]} p_j^\star \cdot \Big(\sum_{t\in[T]} \mu_j^t - \sum_{t\in[T]} \mathbb{E}[\langle \mathbf{p}^t, \mathbf{g}^t\rangle] \Big) \,.$$

Proof First, recall by the first-order definition of of convexity that for any $\mathbf{p}^t \in \Delta_m$:

$$f(\mathbf{p}^\star) \;\geq\; f(\mathbf{p}^t) + \big\langle \nabla f(\mathbf{p}^t), \mathbf{p}^\star - \mathbf{p}^t \big\rangle \,.$$

Rearranging and summing over all t gives

$$\sum_{t\in[T]} f(\mathbf{p}^t) - f(\mathbf{p}^\star) \;\leq\; \sum_{t\in[T]} \big\langle \nabla f(\mathbf{p}^t), \mathbf{p}^t - \mathbf{p}^\star \big\rangle \,.$$

Now taking expectations on both sides, we can write

$$\begin{aligned}
\sum_{t\in[T]} \mathbb{E}[f(\mathbf{p}^t) - f(\mathbf{p}^\star)] \;&\leq\; \sum_{t\in[T]} \mathbb{E}\big[\big\langle \nabla f(\mathbf{p}^t), \mathbf{p}^t - \mathbf{p}^\star \big\rangle\big] \\
&=\; \sum_{t\in[T]} \mathbb{E}\big[\mathbb{E}_{\mathbf{p}^t}\big[\big\langle \nabla f(\mathbf{p}^t), \mathbf{p}^t - \mathbf{p}^\star \big\rangle\big]\big] \\
&=\; \sum_{t\in[T]} \mathbb{E}\big[\big\langle \mathbb{E}_{\mathbf{p}^t}[\nabla f(\mathbf{p}^t)], \mathbf{p}^t - \mathbf{p}^\star \big\rangle\big] \,,
\end{aligned}$$

where we applied the law of iterated expectation. Now recall that under Assumption 1, each reward $\mathbf{g}^t$ if of the form: $\mathbf{g}^t = -(\nabla f(\mathbf{p}^t)/G) + \mathbf{b}^t$, where $\mathbf{b}^t$ is a zero-mean random vector. Thus for every t, it follows that $\mathbb{E}_{\mathbf{p}^t}[\nabla f(\mathbf{p}^t)] = -G \cdot \mathbb{E}[\mathbf{g}^t]$. This allows us to further simplify and write

$$\begin{aligned}
\sum_{t\in[T]} \mathbb{E}[f(\mathbf{p}^t) - f(\mathbf{p}^\star)] \;&\leq\; \sum_{t\in[T]} -G \cdot \mathbb{E}[\langle \mathbf{g}^t, \mathbf{p}^t - \mathbf{p}^\star] \\
&=\; G \sum_{t\in[T]} \mathbb{E}[\langle \mathbf{p}^\star - \mathbf{p}^t, \mathbf{g}^t\rangle] \;=\; G \sum_{t\in[T]} \mathbb{E}[\langle \mathbf{p}^\star, \mathbf{g}^t\rangle] - \sum_{t\in[T]} \mathbb{E}[\langle \mathbf{p}^t, \mathbf{g}^t\rangle] \,.
\end{aligned} \tag{18}$$

Given that $\mathbf{p}^\star$ is fixed, observe for every t that $\mathbb{E}[\langle \mathbf{p}^\star, \mathbf{g}^t\rangle] = \langle \mathbf{p}^\star, \boldsymbol{\mu}^t\rangle = \sum_{j\in[m]} p_j^\star \cdot \mu_j^t$, and substituting this back into (18) gives

$$\begin{aligned}
\sum_{t\in[T]} \mathbb{E}[f(\mathbf{p}^t) - f(\mathbf{p}^\star)] \;&\leq\; G \sum_{t\in[T]} \Big(\sum_{j\in[m]} p_j^\star \cdot \mu_j^t \Big) - \sum_{t\in[T]} \mathbb{E}[\langle \mathbf{p}^t, \mathbf{g}^t\rangle] \\
&=\; G \sum_{j\in[m]} p_j^\star \cdot \Big(\sum_{t\in[T]} \mu_j^t - \sum_{t\in[T]} \mathbb{E}[\langle \mathbf{p}^t, \mathbf{g}^t\rangle] \Big) \,.
\end{aligned} \tag{19}$$

Here, the last line follows from the fact that $\sum_{j\in[m]} p_j^\star = 1$ and that $\mathbb{E}[\langle \mathbf{p}^t, \mathbf{g}^t\rangle]$ has no dependence on j. Now on the other hand, given that f is convex, observe also by Jensen's inequality that

$$\frac{1}{T} \sum_{t\in[T]} f(\mathbf{p}^t) \;\geq\; f\Big(\frac{1}{T} \sum_{t\in[t]} \mathbf{p}^t \Big) \;=\; f(\widetilde{\mathbf{p}}) \,,$$

which holds given that $\widetilde{\mathbf{p}}$ is a convex combination of points. Thus taking expectation, we have

$$\mathbb{E}[f(\widetilde{\mathbf{p}})] - f(\mathbf{p}^\star) \;\leq\; \frac{1}{T}\sum_{t\in[T]} \mathbb{E}[f(\mathbf{p}^t)] - f(\mathbf{p}^\star)\;. \tag{20}$$

Finally, multiplying expression (19) by $\frac{1}{T}$ and combining it with expression (20) yields the statement of the lemma. ■

Proof of Theorem 2.4

Given the sequence $\{\mathbf{p}^t\}$, observe that Lemma G.1 allows us to upper bound the minimization error at the point $\widetilde{\mathbf{p}}$ by a convex combination of the "arm-j regret," i.e., the quantity $\sum_{t\in[T]}\mu_j^t - \sum_{t\in[T]}\mathbb{E}[\langle \mathbf{p}^t, \mathbf{g}^t\rangle]$. Now recall from the points made in Remark F.3 that if we have an initial uniform lower bound on the adoption mass p_j^t at every arm j, then the regret bound from Proposition F.1 can also be used to bound the quantity $\sum_{t\in[T]}\mu_j^t - \sum_{t\in[T]}\mathbb{E}[\langle \mathbf{p}^t, \mathbf{g}^t\rangle]$ for each j.

Nnote that in the context of Assumption 1, we assume that the reward generation sequence $\{\mathbf{g}^t\}$ is adversarial in the sense that the means $\boldsymbol{\mu}^t$ will vary with time. In general, we can thus make only pessimistic uniform assumptions about the adoption mass across all coordinates. For this reason, we require the same set of constraints on T as in Theorem 2.3 for the general, adversarial reward setting (i.e., T can grow at most logarithmically in n). Then we can similarly apply the regret bound from Proposition F.1 with T constrained as in Theorem 2.3, and starting from the uniform distribution $\mathbf{p}^0 = \mathbf{1}/m$.

Thus using a similar calculation as in Theorem 2.3, using the arguments above from Remark F.3, and subject to the constraints on T, we have for each $j \in [m]$:

$$\frac{1}{T}\cdot\sum_{t\in[T]}\mu_j^t - \mathbb{E}[\langle \mathbf{p}^t, \mathbf{g}^t\rangle] \;\leq\; O\left(\sqrt{\frac{\log m}{T}}\right) + \widetilde{O}\left(\frac{\sigma m}{n^\epsilon} + \frac{\sigma m T}{n^c}\right)\;.$$

where $\{\mathbf{p}^t\}$ is the sequence induced using the β-softmax-compare or β-sigmoid-adopt protocols with appropriately tuned β (in particular, the same settings as in Theorem 2.3).

Now because the right hand side of this expression is uniform over all $j \in [m]$, taking a convex combination of this inequality with respect to $\mathbf{p}^\star$, multiplying both sides by G, and applying the reduction from Lemma G.1 yields the statement of Theorem 2.3.