Online Discrete Optimization in Social Networks in the Presence of Knightian Uncertainty[*]

Maxim Raginsky[†]   Angelia Nedić[‡]

First version: July 1, 2013; revision: January 27, 2015

Abstract

We study a model of collective real-time decision-making (or learning) in a social network operating in an uncertain environment, for which no a priori probabilistic model is available. Instead, the environment's impact on the agents in the network is seen through a sequence of cost functions, revealed to the agents in a causal manner only after all the relevant actions are taken. There are two kinds of costs: individual costs incurred by each agent and local-interaction costs incurred by each agent and its neighbors in the social network. Moreover, agents have inertia: each agent has a default mixed strategy that stays fixed regardless of the state of the environment, and must expend effort to deviate from this strategy in order to respond to cost signals coming from the environment. We construct a decentralized strategy, wherein each agent selects its action based only on the costs directly affecting it and on the decisions made by its neighbors in the network. In this setting, we quantify social learning in terms of regret, which is given by the difference between the realized network performance over a given time horizon and the best performance that could have been achieved in hindsight by a fictitious centralized entity with full knowledge of the environment's evolution. We show that our strategy achieves the regret that scales polylogarithmically with the time horizon and polynomially with the number of agents and the maximum number of neighbors of any agent in the social network.

1 Introduction

1.1 Risk vs. uncertainty in social learning and optimization

Decision-making and optimization based on information dispersed among a large number of agents are topics of significant current interest, from both theoretical and practical points of view. Existing literature, which is vast, covers a wide variety of models with different assumptions on the information structure, i.e., who is allowed to observe what, and on the agents' capabilities, i.e., what they are allowed to or able to do with their observations. For example, canonical models of Bayesian learning [1] assume complete and truthful sharing of all relevant information among all agents, who are also endowed

[†]Department of Electrical and Computer Engineering and the Coordinated Science Laboratory, University of Illinois at Urbana–Champaign, Urbana, IL 61801, USA. E-mail: maxim@illinois.edu.

[‡]Department of Industrial and Enterprise Systems Engineering and the Coordinated Science Laboratory, University of Illinois at Urbana–Champaign, Urbana, IL 61801, USA. E-mail: angelia@illinois.edu.

with essentially unlimited computational power. There are also generalizations to noisy signals, but it is typically assumed that the source of noise is nonadversarial. A related framework of Bayesian dynamic games [2, 3] considers sequential decisions by a large collection of agents, where each agent has perfect recall of all decisions made in the past (but not necessarily of all *information* used to arrive at those decisions).

Recently, however, emphasis has shifted towards decision-making in *social networks*, where information sharing is limited to small groups of agents — e.g., when an individual is deciding whether to buy a particular product, she can directly observe similar decisions made by her friends, neighbors or coworkers. Thus, a social network can be modeled as a graph, where each vertex corresponds to an agent while edges correspond to pairwise interactions between agents [4]. Most theoretical studies of social decision-making rest on the following basic framework [1, 4]: (i) there is some unknown parameter associated with the environment in which the network is situated; (ii) each agent receives a private signal stochastically related to this parameter; and (iii) agents select actions by aggregating their private signals with any information they receive from their neighbors in the social network. The main question is whether the agents can *learn* enough about this parameter of interest under the given information structure and the constraints on their information-processing capabilities. For instance, Acemoglu et al. [5] consider Bayesian learning in dynamic social networks with randomly evolving neighborhoods, while Jadbabaie et al. [6] examine a non-Bayesian model of learning in a fixed network, where agents form their beliefs about the underlying parameter by mixing Bayesian updates computed on the basis of their private information with the beliefs of their neighbors.

There are several key modeling assumptions underlying these and similar works:

(S1) The environment is static, meaning that the underlying parameter is drawn from a fixed probability distribution once and for all, and does not change throughout the learning process.

(S2) Each agent has a coherent probabilistic model of the environment in the form of a joint probability measure on the Cartesian product of the parameter space and the agent's private signal space.

(S3) The agents have no intrinsic goals or default strategies unrelated to the state of the environment.

In this paper, we introduce a model of discrete-time decision-making in social networks that departs from all three of these assumptions. In particular, on the descriptive level, our setting has the following features:

(D1) The environment is dynamic, and no agent has a model of its evolution.

(D2) In view of the item above, the environment does not admit a probabilistic representation. Instead, at each time step, each agent receives a signal that quantifies the *costs* of all possible actions that could be taken by this agent and its neighbors in the social network in the current state of the environment.

(D3) No agent is compelled to take only those actions that would entail lower costs. Instead, each agent has a default mixed strategy that stays fixed regardless of the state of the environment, and must expend effort in order to deviate from this strategy.

The distinction between the probabilistic (or Bayesian) view of the environment stipulated in S1–S2 and the nonprobabilistic view laid out in D1–D2 is along the same lines as the distinction between *risk* and *uncertainty* made in 1921 by Frank Knight [7]. According to Knight, risk describes situations with outcomes modeled by random variables with known probability distributions, while uncertainty pertains to

situations in which no such probabilistic description is available or even possible. For instance, uncertainty may arise due to the presence of boundedly rational agents with different sets of values, norms, and abilities. Despite the clear conceptual and practical significance of this distinction, there has been little effort in economics to formalize it mathematically. One of the few exceptions is the work of Bewley [8, 9], who studies the behavior of a decision-making agent interacting in real time with an environment in a state of *punctuated equilibrium* — i.e., intervals of (relative) stability are interrupted by "shocks," corresponding to sharp and unpredictable changes. The Knightian aspect is embodied in the premise that the agent is unable to anticipate the frequency and the nature of these shocks in advance, and so may be caught by surprise. An ideal Bayesian risk-minimizing agent, on the other hand, is not really surprised by anything, since by definition it has already assigned subjective beliefs and utilities to all possible contingencies. Moreover, a Knightian agent may exhibit *inertia*, i.e., a tendency to stick to some default strategy unless there is a sufficiently strong signal from the environment compelling the agent to deviate from the status quo.

Thus, we are interested in the collective decision-making (or learning) capabilities of social networks in the presence of Knightian uncertainty, as captured by the assumptions D1–D3. We quantify learning in terms of *regret*, i.e., the difference between the realized performance of the network over a given time horizon and the best performance that could have been achieved in hindsight by a fictitious centralized entity with full knowledge of the environment's evolution. The performance criterion is induced by a time-varying sequence of composite objective functions that incorporate the total cost of actions taken by all the agents and the total effort expended by the agents in deviating from their individual default strategies.

1.2 A sketch of the model and a summary of results

Let us give a more formal description. We start by considering a *single* agent who must choose an action from a finite set of alternatives, while attempting to balance the instantaneous cost of that action against a desire to minimize effort by sticking to some default (or *status quo*) behavior. Mathematically, we may model such an agent as follows. Let X denote the set of all possible actions, and let μ_0 be a fixed probability distribution on X, where for each $x \in \mathsf{X}$ we interpret $\mu_0(x)$ as the default probability that the agent will choose action x. (For instance, we may imagine a large population of similar agents and take $\mu_0(x)$ as the fraction of agents that tend to choose action x by default.) Without loss of generality, we may suppose that $\mu_0(x) > 0$ for all $x \in \mathsf{X}$. Now let $f \colon \mathsf{X} \to \mathbb{R}$ be a function that prescribes the cost of each action. If we allow the agent to randomize, then a reasonable strategy for the agent would be to choose a random action according to

$$\pi = \underset{\nu \in \mathscr{P}(\mathsf{X})}{\operatorname{argmin}} \left\{ \beta \langle \nu, f \rangle + D(\nu \| \mu_0) \right\},$$

where $\mathscr{P}(\mathsf{X})$ is the space of all probability distributions on X,

$$\langle \nu, f \rangle \triangleq \sum_{x \in \mathsf{X}} \nu(x) f(x)$$

is the expected cost of a random action sampled from the set X according to $\nu \in \mathscr{P}(\mathsf{X})$,

$$D(\nu \| \mu_0) \triangleq \sum_{x \in \mathsf{X}} \nu(x) \ln \frac{\nu(x)}{\mu_0(x)}$$

is the *relative entropy* (or *Kullback–Leibler divergence*[1]) between v and μ_0 [10], and $\beta > 0$ is a parameter that controls the trade-off between loss aversion (i.e., the desire to minimize expected cost) and inertia (i.e., the desire to stick to default behavior) of a Knightian decision-maker [8]. Thus, the addition of the relative-entropy term is an analytically tractable means of penalizing excessive deviations from the default strategy. A simple argument based on the method of Lagrange multipliers gives an explicit form of the solution π:

$$\pi(x) = \frac{\mu_0(x) \exp\left(-\beta f(x)\right)}{Z(\beta)}, \tag{1}$$

where $Z(\beta) = \langle \mu_0, \exp(-\beta f) \rangle$ is a normalization factor. This strategy is well-known in econometrics under the name of *multinomial logit choice model* [11], and it also plays a prominent role in the context of distributionally robust optimization [12]. Probability distributions of this form are also well-known in statistical physics under the name of *Gibbs measures* (see Section 1.4 for more details), where f plays the role of an energy function and β is the *inverse temperature*. Note, in particular, the two extreme regimes: when $\beta = 0$ (infinite temperature), the cost f has no influence on the agent, and we have $\pi = \mu_0$; on the other hand, as $\beta \to \infty$ (zero temperature), the agent has no inertia, and π will converge to the uniform distribution supported on the set of minimizers of f.

In the above formulation, the agent knows the cost function f and thus has no uncertainty about the consequences of various actions; we have only captured the agent's inertia by means of the relative-entropy term. Now, let us bring in an element of time and consider a boundedly rational agent operating in a dynamic environment. Bounded rationality comes from the fact that the agent is unable (or unwilling) to construct an intelligible model of its environment, in the spirit of Knightian uncertainty. The agent must take a sequence of random actions $X_1, \ldots, X_T \in \mathsf{X}$ at discrete time steps $t = 1, 2, \ldots, T$. We suppose also that, at each time t, the costs of each action change unpredictably, and the agent only finds out the current cost function $f_t : \mathsf{X} \to \mathbb{R}$ after having taken the action X_t. However, the agent keeps track of all past cost functions $f_1, \ldots, f_{t-1}$, and may use this information when choosing X_t. We assume that the environment is *nonreactive*, i.e., the sequence $f_1, \ldots, f_T$ of instantaneous cost functions is fixed in advance. Finally, we assume that the default distribution μ_0 over the action set X does not change.

More formally, let $\pi_t \in \mathscr{P}(\mathsf{X})$ denote the distribution of X_t chosen by the agent based on all available information at time t. Then, the *instantaneous loss* incurred by the agent at time t is given by

$$\ell_t(\pi_t) \triangleq \beta \langle \pi_t, f_t \rangle + D(\pi_t \| \mu_0). \tag{2}$$

Due to the agent's limited forecasting ability, we adopt a backward-looking optimality criterion based on *worst-case regret*: If the cost functions f_t are chosen from some fixed class $\mathscr{F}$ known to the agent, the agent should choose a strategy (i.e., a rule for mapping all available information at each time t to a probability distribution π_t of X_t) so as to minimize the worst-case regret

$$R_T(\mathscr{F}) \triangleq \sup_{f_1, \ldots, f_T \in \mathscr{F}} R_T(f^T), \tag{3}$$

where

$$R_T(f^T) \triangleq \sum_{t=1}^{T} \ell_t(\pi_t) - \inf_{v \in \mathscr{P}(\mathsf{X})} \sum_{t=1}^{T} \ell_t(v) \tag{4}$$

---

[1]The Kullback–Leibler divergence is a commonly used measure of (dis)similarity between probability distributions; we discuss some of its salient properties in Section 1.4.

is the regret with respect to a fixed sequence $f^T = (f_1, \ldots, f_T)$ of instantaneous costs. The regret quantifies the worst-case gap between the cumulative loss after T time steps and the smallest cumulative loss that could have been achieved in hindsight had the agent been aware of the entire sequence $f^T = (f_1, \ldots, f_T)$ of instantaneous costs ahead of time. Indeed, if we normalize both sides of (4) by βT, then the minimum of the per-round regret

$$\bar{R}_{\beta,T}(\mathscr{F}) = \sup_{f_1,\ldots,f_T \in \mathscr{F}} \left\{ \frac{1}{T} \left[\sum_{t=1}^{T} \langle \pi_t, f_t \rangle + \frac{1}{\beta} D(\pi_t \| \mu_0) \right] - \inf_{v \in \mathscr{P}(\mathsf{X})} \left[\left\langle v, \frac{1}{T} \sum_{t=1}^{T} f_t \right\rangle + \frac{1}{\beta} D(v \| \mu_0) \right] \right\}. \tag{5}$$

over all strategies for the agent quantifies the smallest gap the agent can secure in the worst case between (a) the average of the expected losses at each round without any knowledge of what will happen in future rounds and (b) the minimum expected loss the agent could attain in a single round against the *empirical average* $(1/T) \sum_{t=1}^{T} f_t$ of the instantaneous costs. Since the agent does not make any probabilistic assumptions about the evolution of the cost sequence $f_1, \ldots, f_T$, the empirical average of this sequence is a reasonable proxy for the "typical" behavior of the environment. Moreover, from the results of Abernethy et al. [13] it follows that, under certain mild conditions allowing the use of the minimax theorem, the per-round minimax regret admits the following equivalent characterization:

$$\bar{R}_{\beta,T}^{*}(\mathscr{F}) = \inf_{\text{strategies}} \bar{R}_{\beta,T}(\mathscr{F})$$

$$= \frac{1}{T} \sup_{\tilde{\mu} \in \mathscr{P}(\mathscr{F}^T)} \mathbb{E}_{f^T \sim \tilde{\mu}} \left\{ \sum_{t=1}^{T} \inf_{v \in \mathscr{P}(\mathsf{X})} \mathbb{E}\left[\langle v, f_t \rangle + \frac{1}{\beta} D(v \| \mu_0) \Big| f^{t-1} \right] - \inf_{v \in \mathscr{P}(\mathsf{X})} \sum_{t=1}^{T} \left[\langle v, f_t \rangle + \frac{1}{\beta} D(v \| \mu_0) \right] \right\}, \tag{6}$$

where $\mathscr{P}(\mathscr{F}^T)$ is the space of all probability measures on T-tuples over $\mathscr{F}$ (with respect to a suitable σ-algebra). This characterization shows that regret minimization is a sequential (or dynamic) generalization of robust Bayesian optimization [12] that takes into account the fact that the agent does accumulate some information about the environment and may use it to some extent to compensate for future uncertainty about the instantaneous cost functions. Indeed, for a fixed $\tilde{\mu} \in \mathscr{P}(\mathscr{F}^T)$, the term corresponding to time t in the first summation in (6) corresponds to the selection of the best strategy π_t when the next instantaneous cost f_t would be drawn from the posterior distribution $\tilde{\mu}(\cdot | f^{t-1})$.

Online decision or prediction problems of this sort have received a great deal of attention in such fields as machine learning, operations research, and finance [14–17]. Their origins date back to a seminal paper of Hannan [18], who has shown that an agent making repeated decisions in a dynamic and uncertain environment will eventually "learn" to act almost as well as if it were aware of the sequence of environment states before beginning to act. To fix ideas, us give an illustrative example in the context of discrete optimization: the *online shortest path* problem [19,20]. Let $G = (V, E)$ be a directed acyclic graph with two distinguished vertices a and b. Let the action space X be the set of all (directed) paths from a to b; it can be identified with a subset of $\{0, 1\}^E$, each of whose elements is a tuple $x = (x_e)_{e \in E}$, such that $x_e = 0$ or 1 depending on whether the edge e is included in the path. The amount of traffic on each edge $e \in E$ varies with time arbitrarily. If we denote the traffic on edge e at time t by $d_{t,e}$, then the total traffic along a given path $x \in \mathsf{X}$ is given by

$$f_t(x) = \sum_{e \in E} d_{t,e} x_e. \tag{7}$$

At each time t, the agent picks a probability distribution π_t over paths from a to b, takes a random path

$X_t \sim \pi_t$, and experiences the average traffic of

$$\langle \pi_t, f_t \rangle = \sum_{x \in X} \pi_t(x) \sum_{e \in E} d_{t,e} x_e.$$

Let $\mathscr{F}$ be the class of all functions of the form (7). Let us first consider the case of no inertia ($\beta \to \infty$). Then it can be shown (see, e.g., [19] or [15, Sec. 5.4]) that the optimal strategy at time t is given by

$$\pi_t(x) \propto \exp\left(-\eta \sum_{s=1}^{t-1} \sum_{e \in E} d_{s,e} x_e\right),$$

where the parameter η is given by $\text{const} \cdot \sqrt{\frac{\ln|X|}{T}}$, resulting in $O\left(\sqrt{T \ln|X|}\right)$ total regret or $O\left(\sqrt{\ln|X|/T}\right)$ per-round regret. Note the intuitive structure of this strategy: it favors (i.e., assigns higher probabilities to) the paths consisting of edges that have experienced the smallest total traffic before time t. Moreover, because of the additive structure of the costs in (7), the per-round computational cost of implementing such a strategy is $O(|E|)$ [19]. The per-round regret has an appealing interpretation as the difference between the total per-round traffic experienced by the agent over T rounds and the average traffic along the best single path the agent could have chosen in hindsight knowing the average amount of traffic $(1/T)\sum_{t=1}^{T} d_{t,e}$ on each edge $e \in E$. If the agent has a finite inertia parameter $\beta > 0$ and a default distribution μ_0 over paths from a to b, then the optimal strategy at time t would take the form

$$\pi_t(x) \propto \mu_0(x) \exp\left(-\frac{\beta}{t} \sum_{s=1}^{t-1} \sum_{e \in E} d_{s,e} x_e\right)$$

(cf. Eq. (1) and Section 3), reflecting the tension between the desire to stick to μ_0 and the effort needed to deviate from it in order to experience less traffic.

Our main interest in this work is in the setting of online decision-making by a *social network* consisting of n agents. This setting has the following salient characteristics:

(i) Each agent takes actions in a finite *base action space* $\{1, \dots, q\}$, so the action space X of the entire network is a Cartesian product $\{1, \dots, q\}^n$. Both the number of alternatives q and the number of agents n are potentially very large.

(ii) The cost functions $f_t \in \mathscr{F}$ decompose into sums of one- and two-variable "local" terms, where each q-ary variable is associated with a separate agent. Thus, when each agent chooses an action, this action affects not only this agent, but also its neighbors in the social network.

(iii) Each of the n agents receives only local information both from other agents and from the environment.

In a large network with local communication, there are two sources of uncertainty for each agent at each time t: uncertainty about the future costs, as well as uncertainty about the actions of all agents outside of that agent's neighborhood in the social network.

Our main contribution is a construction of a decentralized strategy that takes into account these features and whose regret is sublinear in the time horizon T and polylogarithmic in the network parameters (the number of agents and the maximum neighborhood size). We first present a *centralized* strategy and analyze its regret (3) with respect to a class $\mathscr{F}$ of cost functions that decompose into individual (per-agent) and pairwise costs, where the latter affect only those agents that are neighbors in

6

the social network. We then develop an approximate decentralized implementation of this centralized strategy using ideas from statistical physics (specifically, the well-known *Glauber dynamics* or the *Gibbs sampler* [21–23]). It should be pointed out that decentralized strategies based on the Glauber dynamics have been studied in the literature on economics [24–26] and on evolutionary dynamics [27] in the context of convergence to equilibrium in large systems consisting of interconnected agents with local interactions. Our main result (Theorem 1) states that, under a certain regularity condition involving the inverse temperature (or inertia) parameter β and the maximum degree of the social network, the regret of the decentralized strategy based on the Glauber dynamics also exhibits favorable scaling as a function of T and network parameters. The proof of Theorem 1 relies on the aforementioned ideas from statistical physics, as well as on some recent developments in the theory of Markov chains — specifically, on Ollivier's notion of a *positive Ricci curvature* of a Markov chain on a complete separable metric space [28].

1.3 Related literature

The most closely related work to ours is a recent paper by Gamarnik et al. [29], which studies decentralized combinatorial optimization of a random locally decomposable objective function and shows that, under a certain correlation decay condition similar to Dobrushin's uniqueness condition from statistical physics (see, e.g., [30, Section V.1]), it is possible to construct polynomial-time approximation schemes relying only on local information. However, that work is concerned exclusively with *static* (or offline) optimization problems, in which the objective function is fixed. On the other hand, just as in [29], we assume that the instantaneous network costs $f_1, \ldots, f_T$ decompose into a sum of individual and pairwise interaction terms, and at each time step each agent is informed only about its own cost and the pairwise costs in its immediate neighborhood. Another difference with [29] is that they allow communication not only between any pair of immediate neighbors in the graph, but also between agents connected by paths of a given length $r \geq 1$. By contrast, we follow the rest of the social network literature and allow direct communication only between neighbors; however, there are also indirect information paths that affect the scaling of the regret. Finally, the connection between the correlation decay conditions in [29] and statistical physics is primarily qualitative, whereas our regularity condition stated in Theorem 1, as well as the technique used in the proof of the theorem, are more directly related to ideas from statistical physics.

There is also extensive literature on regret minimization in multiagent games, e.g., [31–34], and in particular in graphical games [35] (a class of games, in which the payoff structure is aligned with the social network governing the agents' interactions). However, in this line of work, regret minimization is a goal of each individual agent, who views the rest of the network as a potential opponent. A typical result is that, provided each agent follows a suitable regret-minimizing strategy, the empirical distribution of the actions converges to some equilibrium (e.g., Nash or correlated equilibrium) of the game. By contrast, we view the social network as a *team* that has a common opponent, the environment. Thus, our work can be viewed as an extension of the classical Bayesian economic theory of teams [36, 37] to the realm of online decision-making in the presence of Knightian uncertainty.

1.4 Some notation and preliminaries

Here, we provide some basic concepts and results that will be used later in the development. The *total variation distance* between any two distributions $\mu, \nu \in \mathscr{P}(\mathsf{X})$ is given by

$$\|\mu - \nu\|_{\mathrm{TV}} \triangleq \frac{1}{2} \sum_{x \in \mathsf{X}} |\mu(x) - \nu(x)|.$$

The *Kullback–Leibler divergence* (or *relative entropy*) between μ and ν is

$$D(\mu\|\nu) = \begin{cases} \left\langle \mu, \ln\dfrac{\mu}{\nu} \right\rangle & \text{if } \operatorname{supp}(\mu) \subseteq \operatorname{supp}(\nu), \\ +\infty & \text{otherwise,} \end{cases}$$

where $\operatorname{supp}(\cdot)$ denotes the support of a probability distribution. The Kullback–Leibler divergence is nonnegative, i.e., $D(\mu\|\nu) \geq 0$ for all $\mu, \nu \in \mathscr{P}(\mathsf{X})$, and positive definite, i.e., $D(\mu\|\nu) = 0$ if and only if $\mu = \nu$. (There are other properties, such as convexity, which we do not use in this paper. The reader is invited to consult any text on information theory, such as Cover and Thomas [10], for more details.) These two quantities are related via the Csiszár–Kemperman–Kullback–Pinsker (CKKP) inequality [10, Lemma 17.3.2][2]

$$\|\mu - \nu\|_{\mathrm{TV}} \leq \sqrt{\frac{1}{2} D(\mu\|\nu)}. \tag{8}$$

We will also need some concepts from statistical physics (see, e.g., [30]). Any probability distribution $\mu \in \mathscr{P}(\mathsf{X})$ defines a family of *Gibbs distributions* indexed by functions $g : \mathsf{X} \to \mathbb{R}$:

$$\mu_g(x) \triangleq \frac{\mu(x) \exp\big(g(x)\big)}{\langle \mu, \exp(g) \rangle}, \qquad \forall x \in \mathsf{X}. \tag{9}$$

In statistical physics, each $x \in \mathsf{X}$ is associated with a possible configuration of a physical system, and $g : \mathsf{X} \to \mathbb{R}$ is the negative energy function. In that context, Eq. (9) describes the probabilities of different configurations when the system with energy function g is in a state of equilibrium with a thermal environment at unit absolute temperature [30]. The following lemma (see, e.g., [30, Lemma V.1.4] for a slightly looser bound) provides some properties of Gibbs distributions that will be useful later on in the development of our main results:

Lemma 1. *Let g, h be any two real-valued functions on X. Then we have*

$$D(\mu_g\|\mu_h) \leq \frac{\|g - h\|_{\mathrm{s}}^2}{8}, \tag{10}$$

and

$$\|\mu_g - \mu_h\|_{\mathrm{TV}} \leq \frac{\|g - h\|_{\mathrm{s}}}{4}, \tag{11}$$

where $\|f\|_s$ is the span seminorm *(or* oscillation*) of a function $f : \mathsf{X} \to \mathbb{R}$ given by*

$$\|f\|_{\mathrm{s}} \triangleq \max_{x \in \mathsf{X}} f(x) - \min_{x \in \mathsf{X}} f(x).$$

The proof is elementary, so we give it in Appendix A for completeness.

---

[2]The inequality (8) is often referred to as simply Pinsker's inequality with reference to the book [38], which was a translation of the original Russian text from 1960. However, in [38] Pinsker established a different bound that can be used to deduce (8), but with a much larger constant in front of the relative entropy on the right-hand side. The tight bound (8) was obtained contemporaneously by Csiszár [39], Kemperman [40], and Kullback [41, 42]. The authors would like to thank Prof. Sergio Verdú for pointing out the correct attribution.

2 The model and problem formulation

We model the social network by a simple undirected graph $G = (V, E)$, where each vertex $v \in V$ is associated with an agent, and the edges $\{u, v\} \in E$ indicate symmetric pairwise interactions (in particular, information exchange) among agents. For each v, we denote by $\partial v \triangleq \{u \in V : \{u, v\} \in E\}$ the set of *neighbors* of v, and let $\partial_+ v \triangleq \{v\} \cup \partial v$ denote the set consisting of agent v and all of its neighbors. The *maximum degree* of G is

$$\Delta \triangleq \max_{v \in V} |\partial v|.$$

Each agent takes actions in the base action set $\{1, \ldots, q\}$. The elements of the set

$$\mathsf{X} \triangleq \left\{ x = (x_v)_{v \in V} : x_v \in \{1, \ldots, q\} \right\}$$

will be referred to as *network action profiles*. For each $v \in V$, we fix a probability measure $\mu_{v,0}$ on $\{1, \ldots, q\}$, and let $\mu_0 \in \mathscr{P}(\mathsf{X})$ denote the product measure

$$\mu_0(x) \triangleq \prod_{v \in V} \mu_{v,0}(x_v), \qquad \text{for all } x \in \mathsf{X}. \tag{12}$$

We assume that each $\mu_{v,0}$ charges every action $a \in \{1, \ldots, q\}$, i.e., $\mu_{v,0}(a) > 0$ for all a. The probability measure $\mu_{v,0}$ describes the default individual behavior of agent v. Finally, we are given two classes of local cost functions: the class Φ of one-variable (vertex) costs $\phi : \{1, \ldots, q\} \to \mathbb{R}$ and the class Ψ of two-variable (edge) costs $\psi : \{1, \ldots, q\} \times \{1, \ldots, q\} \to \mathbb{R}$. With this, we denote by $\mathscr{F} = \mathscr{F}_{\Phi,\Psi}$ the space of all functions $f : \mathsf{X} \to \mathbb{R}$ of the form

$$f(x) = \sum_{v \in V} \phi_v(x_v) + \sum_{\{u,v\} \in E} \psi_{u,v}(x_u, x_v), \tag{13}$$

where $\phi_v \in \Phi$ and $\psi_{u,v} \in \Psi$ for all $v \in V$ and all $\{u, v\} \in E$.

The interaction among the agents and the environment takes place according to the following protocol: Initially, each agent $v \in V$ starts out with an empty information set $I_{v,0} = \varnothing$ and draws an action $X_{v,0} \in \{1, \ldots, q\}$ at random according to $\mu_{v,0}$, independently of all other agents. At each discrete time step $t \in \{1, \ldots, T\}$, a single agent $U_t \in V$ is activated uniformly at random independently of all other past data. This agent takes a random action $X_{U_t,t}$ on the basis of all information currently available to it, while all other agents $v \in V \setminus \{U_t\}$ replay their actions from the previous time step $t - 1$. Once the network action profile $X_t = (X_{v,t} : v \in V)$ for time t is generated, each agent v observes its instantaneous cost function $\phi_{v,t}$, its instantaneous local-interaction cost functions $\psi_{u,v,t}$ for $u \in \partial v$, and the decisions of all its neighbors (of course, the agent knows its own decision $X_{v,t}$). Formally, each agent $v \in V$ at time t observes $\iota_{v,t}$, where

$$\iota_{v,t} = \left(\phi_{v,t}; (\psi_{u,v,t} : u \in \partial v); (X_{u,t} : u \in \partial_+ v) \right), \tag{14}$$

and updates its information to $I_{v,t} = (I_{v,t-1}, \iota_{v,t})$. Here, $(\phi_{v,t})_{v \in V}$ and $(\psi_{u,v,t})_{\{u,v\} \in E}$ are the local costs for each agent and for each pair of interacting agents that the environment has generated for time t. As we mentioned earlier, we assume that the environment is nonreactive, i.e., all the cost functions are fixed in advance but revealed to the agents sequentially. Figure 1 gives a summary of this process.

Parameters: base action set $\{1,\dots,q\}$; network graph $G = (V,E)$; default probability measures $\mu_{v,0}$ for all $v \in V$; local function classes Φ, Ψ; number of rounds $T \in \mathbb{N}$.

Initialization of information sets: for each $v \in V$, let $I_{v,0} = \varnothing$.

Initialization of actions: for each $v \in V$, draw $X_{v,0}$ at random according to $\mu_{v,0}$, independently of all other v's.

For each round $t = 1, 2, \dots, T$:

(1) An agent $U_t \in V$ is chosen uniformly at random.

(2) Agent U_t draws a random action $X_{U_t,t}$ on the basis of its current information $I_{U_t,t-1}$; all other agents $v \in V \backslash \{U_t\}$ replay their most recent action: $X_{v,t} = X_{v,t-1}$.

(3) Each agent $v \in V$ observes

　　– the current cost functions $\phi_{v,t} \in \Phi$ and $\psi_{u,v,t} \in \Psi$ for all $u \in \partial v$

　　– the actions $X_{u,t}$ for all $v \in \partial_+ v$,

and updates its information set to $I_{v,t} = (I_{v,t-1}, \iota_{v,t})$, where $\iota_{v,t}$ is the new information available to agent at time t (cf. Eq. (14)).

Figure 1: Online discrete optimization in a network of agents with local interactions.

For each $t = 1, \dots, T$, let μ_t denote the probability distribution of the network action profile X_t.[3] For a fixed sequence of cost functions selected by the environment, the probability measures $\mu_1, \dots, \mu_T$ are fully specified given the initial condition μ_0 in (12) and the sequence of conditional probability distributions

$$\mathbb{P}_{t+1}(x_{t+1}|I_t) = \frac{1}{|V|} \sum_{v \in V} \mathbb{P}_{v,t+1}(x_{v,t+1}|I_{v,t}) \mathbf{1}_{\{x_{-v,t+1} = x_{-v,t}\}}, \qquad t = 0, 1, \dots, T-1 \tag{15}$$

where $\mathbf{1}_{\{\cdot\}}$ is an indicator function that takes value 1 when the logical predicate $\{\cdot\}$ is true and is 0 otherwise, $I_t = (I_{v,t} : v \in V)$ is all the information available immediately after time t, $\mathbb{P}_{v,t+1}(\cdot|I_{v,t})$ is the conditional distribution (or local stochastic update rule) according to which agent v draws its action $x_{v,t+1}$, while $x_{-v,t}$ is the $(|V| - 1)$-tuple obtained from x_t by deleting the coordinate corresponding to agent v, i.e., $x_{-v,t} \triangleq (x_{u,t} : u \in V \backslash \{v\})$. The instantaneous loss incurred by the network at time t is given by

$$\ell_t(\mu_t) = \beta \langle \mu_t, f_t \rangle + D(\mu_t \| \mu_0),$$

where

$$f_t(x) = \sum_{v \in V} \phi_{v,t}(x_v) + \sum_{\{u,v\} \in E} \psi_{u,v,t}(x_u, x_v)$$

is the instantaneous cost function for the entire network at time t. After T rounds, the *regret* of the network with respect to the sequence $f_1, \dots, f_T$ is

$$R_T^{\mathrm{LI}}(f^T) \triangleq \sum_{t=1}^T \ell_t(\mu_t) - \inf_{v \in \mathscr{P}(\mathsf{X})} \sum_{t=1}^T \ell_t(v)$$

---

[3]We will adhere to the following convention: we will use μ_t (respectively, π_t) to denote the distribution of the action profile X_t in the decentralized (respectively, centralized) scenario.

where the superscript LI stands for "local interaction." The corresponding worst-case regret is

$$R_T^{\text{LI}}(\mathcal{F}) \triangleq \sup_{f_1,\dots,f_T \in \mathcal{F}} R_T^{\text{LI}}(f^T). \tag{16}$$

Our objective is to design the local stochastic update rules $\mathbb{P}_{v,t}(\cdot|I_{v,t})$ for all $v \in V$ and all $t \in \{1,\dots,T\}$ to guarantee that the regret (16) is sublinear in T and polynomial in the inverse temperature parameter β, the number of basic actions q, the size $|V|$ of the network, and the maximum number Δ of each agent's neighbors in the social graph.

3 The main results

To motivate our design of a decentralized strategy, we start by developing a particular centralized scheme that, as we shall see, can be well-approximated with a natural distributed implementation. Consider a fixed but arbitrary sequence of instantaneous cost functions $f_1,\dots,f_T \in \mathcal{F}$ chosen by the environment. Our centralized strategy is obtained by the following recursive construction. Suppose that the distributions $\pi_1,\dots,\pi_t$ of the action profiles $X_1,\dots,X_t$ have already been chosen. We choose the next π_{t+1} to balance the greedy tendency to minimize the most recent instantaneous loss $\ell_t(\cdot) = \beta\langle\cdot, f_t\rangle + D(\cdot\|\mu_0)$ against the cautious tendency to stay close to what worked well in the past, i.e., π_t. Hence, a good candidate for π_{t+1} is

$$\pi_{t+1} = \underset{\pi \in \mathscr{P}(\mathsf{X})}{\arg\min} \left\{ \gamma_t \left[\beta\langle\pi, f_t\rangle + D(\pi\|\mu_0) \right] + D(\pi\|\pi_t) \right\}, \tag{17}$$

where the weight $\gamma_t > 0$ controls the trade-off between the greedy and the cautious behavior. This construction is reminiscent of the so-called *mirror descent algorithms* for online convex optimization [15, Chapter 11], with γ_t viewed as a step size at time t. In contrast to the mirror descent algorithms, here the optimization is performed in the space of probability measures and there is no linearization of the objective function. An application of the method of Lagrange multipliers leads to the following solution:

$$\pi_1 = \mu_0 \quad \text{and} \quad \pi_{t+1}(x) = \frac{\left(\mu_0^{\gamma_t}(x)\pi_t(x)\exp\left(-\gamma_t\beta f_t(x)\right)\right)^{\frac{1}{1+\gamma_t}}}{Z_{t+1}}, \quad t = 1, 2, \dots \tag{18}$$

where Z_{t+1} is the normalization constant ensuring that π_{t+1} is a bona fide probability distribution. We will work with $\gamma_t = \frac{1}{t}$. We now summarize the key properties of this strategy.

Proposition 1. *The strategy* (18), *with* $\gamma_t = \frac{1}{t}$, *has the following properties:*

  1. *For any* $t = 0, 1, 2, \dots$, *the distribution* π_{t+1} *can be expressed as*

$$\pi_{t+1}(x) = \frac{\mu_0(x)\exp\left(-\beta F_t(x)\right)}{\tilde{Z}_{t+1}}, \tag{19}$$

  where $\tilde{Z}_{t+1}$ *is a normalization constant, and the functions* F_t *are given by*

$$F_t(x) = \begin{cases} 0, & t = 0, \\ \frac{1}{t+1}\sum_{s=1}^t f_s(x), & t \geq 1. \end{cases} \tag{20}$$

2. *Suppose that the functions $\phi \in \Phi$ and $\psi \in \Psi$ take values in the interval $[-1,1]$. Then, for all $t \geq 1$,*

$$D(\pi_t \| \pi_{t+1}) \leq 2 \left(\frac{\beta |V|(\Delta + 1)}{t+1} \right)^2. \tag{21}$$

3. *It achieves the following bound on the worst-case regret: for all $T \geq 1$,*

$$R_T(\mathcal{F}) \leq 2 \left(\beta |V|(\Delta + 1) \right)^2 \ln(T + 1) + \ln \frac{1}{\theta}, \tag{22}$$

where $\mathcal{F} = \mathcal{F}_{\Phi, \Psi}$ and

$$\theta \triangleq \min_{a \in \{1, \ldots, q\}} \mu_0(a).$$

Several observations and remarks are in order:

1. The $O(\ln T)$ scaling of the regret is a consequence of the general fact that, due to the presence of the relative entropy term, the loss functions ℓ_t are *strongly convex* with respect to the total variation norm (see, e.g., [43]). In fact, there are matching lower bounds [44] that show that this scaling of the regret is optimal for strongly convex losses. However, our analysis of the centralized strategy in (18) is self-contained, and the three parts of Proposition 1 reflect the logical structure of the proof: first, we show that the strategy at time $t + 1$ is given by a Gibbs measure determined by the empirical average of the instantaneous costs revealed up to time t; then we show that the difference between the strategies at successive times t and $t+1$, as measured by the relative entropy $D(\pi_t \| \pi_{t+1})$, decays rapidly with t; and finally we use these intermediate results to derive a bound on the overall regret. Thus, even though the result in Proposition 1 is not new, it differs from the rest of the literature by its emphasis on the *dynamical properties* of the strategy in (18), which we will exploit in the proof of our main result.

2. Implementation of the above centralized strategy does not require advance knowledge of the time horizon T.

3. Inspection of the non-recursive form (19) of π_{t+1} sheds light on the role of the decaying factor $1/t$ in (17): it is used to *dampen* the influence of past instantaneous costs $f_1, \ldots, f_t$. In particular, at time t, each cost function enters into the strategy π_t with the same weight $1/(t + 1)$. As we will see later, this averaging is crucial in ensuring that we can approximate each global randomized strategy π_t using purely local update rules.

4. From the standpoint of the influence of the initial distributions $\mu_{v,0}$, the regret bound in (22) is minimized when $\mu_{v,0}$ is the uniform distribution for all $v \in V$, resulting in

$$R_T(\mathcal{F}) \leq 2 \left(\beta |V|(\Delta + 1) \right)^2 \ln(T + 1) + |V| \ln q.$$

5. The fact that the regret is proportional to $(|V|(\Delta + 1))^2$ is not surprising in light of the fact that each cost function f_t is a sum of $|V|(\Delta + 1)$ terms, each of which is bounded by 1. Thus, $\|f_t\|_s = O(|V|(\Delta + 1))$. Since the regret is governed by the relative-entropy drift terms $D(\pi_t \| \pi_{t+1})$, by Lemma 1 we expect it to scale with $\max_t \|f_t\|_s = O((|V|(\Delta + 1)^2)$.

We now use this centralized scheme to construct appropriate local update rules $\mathbb{P}_{v,t+1}(\cdot|I_{v,t})$ for all $v \in V$ and $t \in \{1,\ldots,T-1\}$. For any cost function f of the form (13), any $v \in V$, and any *boundary condition* $x_{\partial v} \in \{1,\ldots,q\}^{|\partial v|}$, we define the local cost at v by

$$f_v(a, x_{\partial v}) \triangleq \phi_v(a) + \sum_{u \in \partial v} \psi_{u,v}(x_u, a), \qquad \forall a \in \{1,\ldots,q\}.$$

Similar notation will be used for time-indexed instantaneous and discounted cumulative costs, i.e., $f_{v,t}$ based on f_t, and $F_{v,t}$ based on F_t. For each $v \in V$ and each t, let

$$\mathbb{P}_{v,t+1}(x_{v,t+1}|I_{v,t}) \triangleq \frac{\mu_{v,0}(x_{v,t+1})\exp\left(-\beta F_{v,t}(x_{v,t+1}, x_{\partial v,t})\right)}{Z_{t+1}(x_{\partial v,t})}, \tag{23}$$

where

$$F_{v,t} = \frac{1}{t+1}\sum_{s=1}^{t} f_{v,s}; \tag{24}$$

The normalization constant $Z_{t+1}(x_{\partial v,t})$ in (23) now depends on the action profile $x_{\partial v,t}$ of the neighborhood of agent v after time t:

$$Z_{t+1}(x_{\partial v,t}) = \sum_{a \in \{1,\ldots,q\}} \mu_{v,0}(a)\exp\left(-\beta F_{v,t}(a, x_{\partial v,t})\right).$$

Note that the history of previous local action profiles $x_{\partial v,1},\ldots,x_{\partial v,t}$ enters into the conditional probabilities (23) only through the most recent action profile $x_{\partial v,t}$. Moreover, if we consider a fixed but arbitrary sequence of network costs $f_1,\ldots,f_T$, then we may simplify our notation by suppressing the dependence of the transition probabilities $\mathbb{P}_{v,t+1}(\cdot|I_{v,t})$ and $\mathbb{P}_{t+1}(\cdot|I_t)$ on the costs $f_1,\ldots,f_t$ and past action profiles $x_1,\ldots,x_{t-1}$. Thus, instead of $\mathbb{P}_{v,t+1}(x_{t+1}|I_{v,t})$, we will write

$$\mathbb{P}_{t+1}(x_{t+1}|x_t) = \frac{1}{|V|}\sum_{v \in V} \mathbb{P}_{v,t+1}(x_{v,t+1}|x_{\partial v,t})\mathbf{1}_{\{x_{-v,t+1}=x_{-v,t}\}}, \tag{25}$$

where we have also used the same convention for the local update rules $\mathbb{P}_{v,t+1}(\cdot|I_{v,t})$. So, when the instantaneous costs $f_1,\ldots,f_T$ are fixed, the network action profiles $X_0, X_1,\ldots,X_T$ form a Markov chain with initial distribution μ_0 and time-inhomogeneous transition probabilities

$$\Pr(X_{t+1} = y|X_t = x) = \mathbb{P}_{t+1}(y|x).$$

One can recognize the Markov transition kernel $\mathbb{P}_{t+1}(x_{t+1}|x_t)$ constructed from (23) according to (15) as one step of the *Glauber dynamics* (or the *Gibbs sampler*) [21–23] induced by the Gibbs distribution π_{t+1} in (19). Consequently, for each t we have the detailed balance (or time-reversibility) property

$$\pi_t(x)\mathbb{P}_t(y|x) = \pi_t(y)\mathbb{P}_t(x|y), \qquad \forall x, y \in \mathsf{X} \tag{26}$$

which implies that π_t is an invariant distribution of $\mathbb{P}_t$ (we give a self-contained proof of this fact in Appendix B). In mathematical economics and game theory, the Glauber dynamics was used by Blume [24] (under the name "log-linear learning") and by Young [25] (under the name "spatial adaptive play") to model the emergence of optimal global behavior in networks of agents with local interactions; see also a recent paper by Alós-Ferrer and Netzer [26] for a discussion of a more general class of *logit-response dynamics* that includes the Glauber dynamics as a special case.

We now state our main result: a bound on the regret of the decentralized strategy based on the Glauber dynamics.

Theorem 1. *Suppose that all functions in* Φ *and* Ψ *take values in* $[-1,1]$. *Also, suppose that the parameter* $\beta > 0$ *and satisfies the following condition:*

$$\Delta\beta < 1. \tag{27}$$

Then the strategy (23)–(25) *based on the Glauber dynamics attains the following worst-case regret:*

$$R_T^{\mathrm{LI}}(\mathscr{F}) \leq R_T^{\mathrm{LI}}(f^T) \leq \frac{|V|}{1 - \Delta\beta}\left(2\beta^2|V|^3(\Delta+1)^2 + K\left(|V|\ln\frac{q^2}{\theta_d} + \ln T\right)\right)\ln(T+1)$$

$$+ 2\left(\beta|V|(\Delta+1)\right)^2\ln(T+1) + T_1|V|\ln\frac{q^2}{\theta_d} + \frac{2\beta|V|^3(\Delta+1)}{1 - \Delta\beta} + \ln\frac{1}{\theta}, \tag{28}$$

where $K \triangleq \max\left\{|V|, \beta|V|^2(\Delta+1)\right\}$,

$$\theta_d \triangleq \min_{v \in V}\ \min_{a \in \{1,\dots,q\}} \mu_{v,0}(a),$$

and $T_1 \equiv T_1(\beta, \Delta, |V|)$ *is a positive constant independent of* T.

A few comments on the interpretation of the above result:

1. The regularity condition in (27) is needed to ensure that the sequence of action profile distributions $\mu_1, \dots, \mu_T$ induced by the Glauber dynamics (23)–(25) closely tracks its centralized counterpart $\pi_1, \dots, \pi_T$ from Proposition 1. It is, essentially, a Dobrushin uniqueness condition from statistical physics (see, e.g., [30, Section V.1]), which is typically used to establish rapid mixing (i.e., convergence to the invariant distribution) of the Glauber dynamics [45, 46]. The correlation decay conditions driving the results of Gamarnik et al. [29] are similar in spirit.

2. We note that β is a fixed exogenous parameter that quantifies the responsiveness of agents to changes in their environment (as reflected through the time-varying instantaneous costs). The regularity condition in (27) therefore involves only the intrinsic parameters of the network and says that the Glauber dynamics (23)–(25) is mixing whenever the temperature parameter $1/\beta$ is larger than the maximum number of neighbors of any agent.

3. Ignoring terms of order lower than $\ln T$, we can express the regret bound more succinctly as

$$R_T^{\mathrm{LI}}(\mathscr{F}) = O\left(\frac{|V|^4(\Delta\beta)^2}{1 - \Delta\beta}\left(\ln\frac{q}{\theta_d}\right)(\ln T)^2\right). \tag{29}$$

Compared to the regret bound (22) in the centralized case, the local-information regret (29) has a worse (but still polynomial) dependence on the network parameters. We also observe that the regret now scales as $(\ln T)^2$, as opposed to the optimal centralized scaling of $\ln T$.

4 Proofs

4.1 Proof of Proposition 1

Part 1. We analyze the strategy (18) with $\gamma_t = \frac{1}{t}$ for all $t \geq 1$. The proof is by induction on t. The base case is $t = 1$, for which, using (18) and the fact that $\gamma_1 = 1$, we have for all $x \in \mathsf{X}$,

$$\pi_2(x) = \frac{\left(\mu_0(x)\pi_1(x)\exp\left(-\beta f_1(x)\right)\right)^{\frac{1}{2}}}{Z_2}.$$

14

Since $\pi_1 = \mu_0$, it follows that

$$\pi_2 = \frac{\mu_0 \exp\left(-\frac{\beta}{2} f_1\right)}{\tilde{Z}_2},$$

thus showing that the expressions in (19) and (20) are valid for $t = 1$ with $\tilde{Z}_2 = Z_2$. Now, suppose that the strategy π_{t+1} satisfies (19) and (20) for a given $t+1$. Then, according to the definition of π_{t+2} via (18), and using the fact that $\gamma_{t+1} = \frac{1}{t+1}$, we have

$$
\begin{aligned}
\pi_{t+2}(x) &= \frac{\left(\mu_0^{\frac{1}{t+1}}(x) \pi_{t+1}(x) \exp\left(-\frac{\beta}{t+1} f_{t+1}(x)\right)\right)^{\frac{t+1}{t+2}}}{Z_{t+2}} \\
&= \frac{\left(\mu_0^{\frac{1}{t+1}}(x) \mu_0(x) \exp\left(-\frac{\beta}{t+1} \sum_{s=1}^{t} f_s(x)\right) \exp\left(-\frac{\beta}{t+1} f_{t+1}(x)\right)\right)^{\frac{t+1}{t+2}}}{\tilde{Z}_{t+1}^{\frac{t+1}{t+2}} Z_{t+2}},
\end{aligned}
$$

where the last equality follows from the induction hypothesis. Hence, for all $x \in \mathsf{X}$, we have

$$\pi_{t+2}(x) = \frac{\mu_0(x) \exp\left(-\frac{\beta}{t+2} \sum_{s=1}^{t+1} f_s(x)\right)}{\tilde{Z}_{t+1}^{\frac{t+1}{t+2}} Z_{t+2}}. \tag{30}$$

Eq. (30) shows that that (19) and (20) hold for $t+2$, with $\tilde{Z}_{t+2} = \tilde{Z}_{t+1}^{\frac{1}{1+\gamma_{t+1}}} Z_{t+2}$. Hence, (19) and (20) are valid for all $t \geq 1$.

Part 2. In order to bound the terms $D(\pi_t \| \pi_{t+1})$, it is convenient to use the form (19) of π_t from Part 1, which expresses π_t as a Gibbs measure. Therefore, we can use Lemma 1 to get

$$D(\pi_t \| \pi_{t+1}) \leq \frac{\beta^2 \left\| F_t - F_{t-1} \right\|_{\mathrm{s}}^2}{8}. \tag{31}$$

Now we need to bound the span seminorm of $F_t - F_{t-1}$. From the definition of F_t in (20) and the relation (30), it can be seen that

$$F_1(x) = \frac{1}{2} f_1(x) \quad \text{and} \quad F_t(x) = \frac{1}{t+1} f_t(x) + \frac{t}{t+1} F_{t-1}(x) \ \text{ for all } t \geq 2. \tag{32}$$

Since $F_0 = 0$, we can write

$$F_t(x) - F_{t-1}(x) = \frac{1}{t+1} \left[f_t(x) - F_{t-1}(x) \right] \quad \text{for all } t \geq 1.$$

Hence,

$$\left\| F_1 - F_0 \right\|_{\mathrm{s}} \leq \frac{1}{2} \| f_1 \|_{\mathrm{s}} \quad \text{and} \quad \left\| F_t - F_{t-1} \right\|_{\mathrm{s}} \leq \frac{1}{t+1} \left(\| f_t \|_{\mathrm{s}} + \left\| F_{t-1} \right\|_{\mathrm{s}} \right) \qquad \text{for all } t \geq 2. \tag{33}$$

Now, using the definition of F_{t-1} and the relation $\| f \|_{\mathrm{s}} \leq 2 \| f \|_\infty$, valid for any function f on X, we have

$$\left\| F_\ell \right\|_{\mathrm{s}} \leq \frac{1}{\ell+1} \sum_{s=1}^{\ell} \| f_s \|_{\mathrm{s}} \leq \frac{2}{\ell+1} \sum_{s=1}^{\ell} \| f_s \|_\infty.$$

By employing Lemma D.1, according to which $\|f_s\|_\infty \le |V|(\Delta+1)$, we further obtain

$$\left\|F_\ell\right\|_s \le 2|V|(\Delta+1) \qquad \text{for all } \ell \ge 1. \tag{34}$$

Using (34) in the expression (33), we get

$$\left\|F_t - F_{t-1}\right\|_s \le \frac{4|V|(\Delta+1)}{t+1}.$$

By substituting the preceding estimate into (31) we obtain, for all $t \ge 1$,

$$D(\pi_t\|\pi_{t+1}) \le \frac{\beta^2}{8}\left(\frac{4|V|(\Delta+1)}{t+1}\right)^2 = 2\left(\frac{\beta|V|(\Delta+1)}{t+1}\right)^2,$$

which gives us (21).

Part 3. We show the bound for the regret $R_T(f^T)$ by first writing down an *exact* expression for it and then developing a suitable upper bound. For every t, we can use the definition (18) of π_{t+1} to write

$$\beta f_t(x) = \ln\mu_0(x) + t\ln\pi_t(x) - (t+1)\left(\ln Z_{t+1} + \ln\pi_{t+1}(x)\right).$$

Therefore, for any $v \in \mathscr{P}(\mathsf{X})$,

$$
\begin{aligned}
\ell_t(v) &= \beta\langle v, f_t\rangle + D(v\|\mu_0)\\
&= \left\langle v, \beta f_t + \ln\frac{v}{\mu_0}\right\rangle\\
&= \left\langle v, \ln\mu_0 + t\ln\pi_t - (t+1)\left(\ln Z_{t+1} + \ln\pi_{t+1}\right) + \ln\frac{v}{\mu_0}\right\rangle\\
&= \left\langle v, t\ln\frac{\pi_t}{\pi_{t+1}} + \ln\frac{v}{\pi_{t+1}}\right\rangle - (t+1)\ln Z_{t+1}\\
&= t\left\langle v, \ln\frac{v}{\pi_{t+1}} - \ln\frac{v}{\pi_t}\right\rangle + \left\langle v, \ln\frac{v}{\pi_{t+1}}\right\rangle - (t+1)\ln Z_{t+1}\\
&= \left[(t+1)D(v\|\pi_{t+1}) - tD(v\|\pi_t)\right] - (t+1)\ln Z_{t+1}.
\end{aligned}
$$

In particular, letting $v = \pi_t$ and using the fact that $D(v\|v) = 0$ for all v, we get

$$\ell_t(\pi_t) = (t+1)\left[D(\pi_t\|\pi_{t+1}) - \ln Z_{t+1}\right].$$

Therefore, summing from $t = 1$ to $t = T$ and telescoping, we obtain

$$
\begin{aligned}
\sum_{t=1}^{T}\left[\ell_t(\pi_t) - \ell_t(v)\right] &= \sum_{t=1}^{T}(t+1)D(\pi_t\|\pi_{t+1}) + \sum_{t=1}^{T}\left[tD(v\|\pi_t) - (t+1)D(v\|\pi_{t+1})\right]\\
&= \sum_{t=1}^{T}(t+1)D(\pi_t\|\pi_{t+1}) + D(v\|\pi_1) - (T+1)D(v\|\pi_{T+1}).
\end{aligned}
$$

Using the fact that $\pi_1 = \mu_0$ and $D(v\|\pi_{T+1}) \ge 0$, we can further bound the regret as follows:

$$R_T(f^T) \le \sum_{t=1}^{T}(t+1)D(\pi_t\|\pi_{t+1}) + \ln\frac{1}{\theta}, \tag{35}$$

16

where we have used the inequality

$$D(\nu\|\mu_0) = \left\langle \nu, \ln\frac{\nu}{\mu_0} \right\rangle \le |V| \ln\frac{1}{\theta} \qquad \text{for any } \nu \in \mathscr{P}(\mathsf{X}),$$

which holds since $\mu_0(x) > 0$ for all $x \in \mathsf{X}$. Upon substituting the bound for $D(\pi_t\|\pi_{t+1})$ from Part 2 into (35), we obtain

$$
\begin{aligned}
R_T(f^T) &\le \sum_{t=1}^{T} 2\,(t+1)\left(\frac{\beta|V|(\Delta+1)}{t+1}\right)^2 + \ln\frac{1}{\theta} \\
&= 2\left(\beta|V|(\Delta+1)\right)^2 \sum_{t=1}^{T}\frac{1}{t} + \ln\frac{1}{\theta} \\
&\le 2\left(\beta|V|(\Delta+1)\right)^2 \ln(T+1) + \ln\frac{1}{\theta},
\end{aligned}
$$

where the last inequality follows from

$$\sum_{t=1}^{T}\frac{1}{t+1} \le \int_{1}^{T+1}\frac{\mathrm{d}t}{t} = \ln(T+1).$$

Since this bound holds uniformly in all $f_1,\dots,f_T \in \mathscr{F}$, we get (22).

4.2 Proof of Theorem 1

Before proceeding with the formal proof, let us briefly outline the intuition behind it. The underlying idea is to express the regret $R_T^{\mathrm{LI}}(f^T)$ for the decentralized local-interaction strategy $\{\mu_t\}_{t=0}^{T}$ as the sum of the regret $R_T(f^T)$ for the centralized strategy $\{\pi_t\}_{t=0}^{T}$ and the extra cost due to decentralization. Theorem 1 provides a bound for $R_T(f^T)$, and the main effort of the proof is in establishing a bound on the total decentralization cost incurred over the time horizon T. In turn, this total decentralization cost depends on the distances between the centralized action profile distribution π_t and its centralized counterpart μ_t for $t = 1,\dots,T$. These distances turn out to be small due to the use of the Glauber dynamics.

More specifically, consider the centralized strategy $\{\pi_t\}_{t=0}^{T+1}$ given in (18). By construction of the local update rules in (23), each "global" probability measure π_t is invariant with respect to the Markov transition kernel $\mathbb{P}_t$ given in (25). Moreover, as the relative entropy bound (21) shows, the probability measures π_t and π_{t+1} are close for every t. Finally, we will show that, under the condition $\Delta\beta < 1$, the conditional distributions $\mathbb{P}_t(\cdot|x)$ and $\mathbb{P}_t(\cdot|y)$ will be close whenever the action profiles x and y are close. As we will demonstrate shortly, these three properties together ensure that, at each time step t, the decentralized action profile distribution $\mu_t = \mathbb{P}_t\mathbb{P}_{t-1}\dots\mathbb{P}_1\mu_0$ will be close to its centralized counterpart π_t. On a "big picture" level, this argument is similar in spirit to the one used by Narayanan and Rakhlin [47] to construct and analyze efficient algorithms for centralized online minimization of a sequence of linear functions on a compact convex subset of a finite-dimensional Euclidean space. However, here we are interested in *decentralized* algorithms for *discrete* optimization. Moreover, the overall proof in [47] is rather technical, drawing on ideas from the Riemannian geometry of interior-point optimization algorithms [48] and random walks on convex bodies [49]. By contrast, our proof is much simpler, and relies on the notion of *positive Ricci curvature* of a Markov chain recently introduced by Ollivier [28] (the reader is invited to consult a recent paper by Joulin and Ollivier [50] for examples of how Ricci curvature ideas can be used to get sharp estimates of convergence rates of MCMC algorithms).

To separate out the key ideas underlying our proof, we have split this section into three parts. The first part (Section 4.2.1) uses the notion of Ricci curvature of Markov chains to obtain uniform error bounds for sampling from a time-varying sequence of probability measures. Because the results of this part may be of independent interest, we formulate them in a much more general setting of complete separable metric spaces. The second part (Section 4.2.2) applies these results to the time-varying Glauber dynamics (23)–(25). Once all the necessary ingredients are in place, we complete the proof of Theorem 1 in the last part (Section 4.2.3).

4.2.1 Positive Ricci curvature and sampling from a time-varying sequence of probability measures

Let (X, ρ) be a complete separable metric space (i.e., a Polish space) equipped with the σ-algebra $\mathscr{B}(X)$ of its Borel subsets. A *Markov transition kernel* on X is a mapping $\mathbb{P}(\cdot|\cdot) : \mathscr{B}(X) \times X \to [0, 1]$, such that (i) $\mathbb{P}(\cdot|x)$ is a probability measure on X for all x and (ii) the mapping $x \mapsto \mathbb{P}(A|x)$ is measurable for every $A \in \mathscr{B}(X)$. We define the action of a Markov kernel $\mathbb{P}$ on a probability measure $\mu \in \mathscr{P}(X)$ as

$$\mathbb{P}\mu(A) \triangleq \int_X \mathbb{P}(A|x)\mu(\mathrm{d}x),$$

and we say that μ is $\mathbb{P}$-*invariant* if $\mu = \mathbb{P}\mu$. The L^1 *Wasserstein distance* (or *transportation distance*) [51] between probability measures $\mu, \nu \in \mathscr{P}(X)$ is defined as

$$W_1(\mu, \nu) \triangleq \inf_{\nu \in C(\mu,\nu)} \int_{X \times X} \rho(x, y)\nu(\mathrm{d}x, \mathrm{d}y), \tag{36}$$

where $C(\mu, \nu)$ denotes the collection of all *couplings* of μ and ν, i.e., all probability measures ν on $X \times X$ with marginals μ and ν. An important *Kantorovich–Rubinstein theorem* (see, e.g., [51, Theorem 1.14]) gives a variational representation of $W_1(\mu, \nu)$:

$$W_1(\mu, \nu) = \sup_{f:\|f\|_{\mathrm{Lip}} \leq 1} \left| \int_X f \mathrm{d}\mu - \int_X f \mathrm{d}\nu \right|, \tag{37}$$

where the supremum is over all real-valued functions f on X with Lipschitz constant

$$\|f\|_{\mathrm{Lip}} \triangleq \sup_{x \neq y} \frac{|f(x) - f(y)|}{\rho(x, y)} \leq 1.$$

Remark 1. When ρ is the *trivial metric*, i.e., $\rho(x, y) = \mathbf{1}_{\{x \neq y\}}$, the Wasserstein distance is equal to the total variation distance: for any $\mu, \nu \in \mathscr{P}(X)$:

$$\|\mu - \nu\|_{\mathrm{TV}} = \inf_{\nu \in C(\mu,\nu)} \int_{X \times X} \mathbf{1}_{\{x \neq y\}}\nu(\mathrm{d}x, \mathrm{d}y) \tag{38}$$

Moreover, for any two $\mu, \nu \in \mathscr{P}(X)$, we can construct the so-called *optimal coupling* $\nu^\star \in C(\mu, \nu)$ that achieves the infimum in (38) (see, e.g., [52, Section 4.2]).

Fix a Markov kernel $\mathbb{P}$ on X. Following Ollivier [28], we say that $\mathbb{P}$ has *positive Ricci curvature* if there exists some $\kappa \in (0, 1]$, such that

$$W_1\big(\mathbb{P}(\cdot|x), \mathbb{P}(\cdot|y)\big) \leq (1 - \kappa)\rho(x, y), \qquad \forall x, y \in X. \tag{39}$$

We will denote the supremum of all such κ by $\mathrm{Ric}(\mathbb{P})$ and call this number the Ricci curvature of $\mathbb{P}$. The following contraction inequality [28, Proposition 20] is key: (39) holds for $\mathbb{P}$ with some $\kappa \in (0,1]$ if and only if

$$W_1(\mathbb{P}\mu, \mathbb{P}\nu) \le (1-\kappa)W_1(\mu, \nu), \qquad \forall \mu, \nu \in \mathscr{P}(\mathsf{X}). \tag{40}$$

We are now ready to develop our main technical tool:

Lemma 2. *Let $\mathbb{P}_1, \mathbb{P}_2, \ldots$ be a sequence of Markov kernels on X with the following properties:*

(i) Each $\mathbb{P}_t$ has a unique invariant distribution π_t and there exists some $\delta_t \in [0,1)$, such that

$$W_1(\pi_t, \pi_{t+1}) \le \delta_t, \qquad t = 1, \ldots, T. \tag{41}$$

(ii) The Ricci curvatures of the $\mathbb{P}_t$'s are uniformly bounded from below by some $\kappa^\star \in (0,1)$:

$$\mathrm{Ric}(\mathbb{P}_t) \ge \kappa^\star, \qquad t = 1, 2, \ldots.$$

Given a probability measure $\mu_1 \in \mathscr{P}(\mathsf{X})$, let $\{\mu_t\}$ be a sequence of probability measures defined recursively via $\mu_{t+1} = \mu_t \mathbb{P}_t$. Then, we have

$$W_1(\mu_t, \pi_t) \le (1-\kappa^\star)^{t-1} W_1(\mu_1, \pi_1) + \mathbf{1}_{\{t \ge 2\}} \sum_{s=1}^{t-1} (1-\kappa^\star)^{t-1-s} \delta_s, \qquad t \ge 1. \tag{42}$$

Proof. By inspection we can see that relation (42) holds for $t = 1$. Next, note that for any $t \ge 0$ we have

$$W_1(\mu_{t+1}, \pi_{t+1}) \le W_1(\mu_{t+1}, \pi_t) + W_1(\pi_t, \pi_{t+1}) \tag{43}$$

$$= W_1(\mathbb{P}_t \mu_t, \mathbb{P}_t \pi_t) + W_1(\pi_t, \pi_{t+1}) \tag{44}$$

$$\le (1-\kappa^\star)W_1(\mu_t, \pi_t) + \delta_t, \tag{45}$$

where (43) is by the triangle inequality, (44) uses the recursive definition of the μ_t's and the $\mathbb{P}_t$-invariance of π_t, and (45) uses the contraction inequality (40) and the assumption (41). Thus, by induction on t, we can find that for all $t \ge 1$,

$$W_1(\mu_{t+1}, \pi_{t+1}) \le (1-\kappa^\star)^t W_1(\mu_1, \pi_1) + \sum_{s=1}^{t} (1-\kappa^\star)^{t-s} \delta_s,$$

which shows (42) for $t \ge 2$. $\qquad\square$

Corollary 1. *Under the assumptions of Lemma 2, for any Lipschitz function $f : \mathsf{X} \to \mathbb{R}$ we have*

$$\left| \int_\mathsf{X} f \mathrm{d}\mu_t - \int_\mathsf{X} f \mathrm{d}\pi_t \right| \le \|f\|_{\mathrm{Lip}} \left((1-\kappa^\star)^{t-1} W_1(\mu_1, \pi_1) + \mathbf{1}_{\{t \ge 2\}} \sum_{s=1}^{t-1} (1-\kappa^\star)^{t-1-s} \delta_s \right), \qquad t = 1, 2 \ldots.$$

Proof. Use (42) and the Kantorovich–Rubinstein formula (37). $\qquad\square$

Remark 2. In the special case when $\delta_t = \delta$ for all t, the bounds of Lemma 2 and Corollary 1 become

$$W_1(\mu_t, \pi_t) \le \frac{\delta}{1-\kappa^\star}$$

and

$$\left| \int_\mathsf{X} f \mathrm{d}\mu_t - \int_\mathsf{X} f \mathrm{d}\pi_t \right| \le \frac{\|f\|_{\mathrm{Lip}} \delta}{1-\kappa^\star}$$

respectively.

4.2.2 Positive Ricci curvature of the time-varying Glauber dynamics

We now particularize these results to our setting, where X is the space of all tuples $x = (x_v : v \in V)$ equipped with the *Hamming distance*

$$\rho_{\mathrm{H}}(x, y) \triangleq \sum_{v \in V} \mathbf{1}_{\{x_v \neq y_v\}}.$$

In this case, the Ricci curvature bounds are equivalent to the so-called *path coupling* bounds of Bubley and Dyer [53] (see also [52, Chapter 14] and [28, Example 17]). In particular, in order to obtain a lower bound on the Ricci curvature of a given Markov kernel $\mathbb{P}$, it suffices to consider only those $x, y \in X$ with $\rho_{\mathrm{H}}(x, y) = 1$. Indeed, suppose that we can find some $\kappa \in (0, 1]$, such that

$$W_1\big(\mathbb{P}(\cdot|x), \mathbb{P}(\cdot|y)\big) \leq 1 - \kappa \qquad \text{for all } x, y \text{ with } \rho_{\mathrm{H}}(x, y) = 1. \tag{46}$$

Then $\mathrm{Ric}(\mathbb{P}) \geq \kappa$. To see this, consider any pair $x, y \in X$ with $\rho_{\mathrm{H}}(x, y) = k$. Then, there exists a sequence $x_1, \ldots, x_{k+1} \in X$, such that $x_1 = x$, $x_{k+1} = y$, and $\rho_{\mathrm{H}}(x_j, x_{j+1}) = 1$ for all $1 \leq j \leq k$. Using this fact, we can write

$$\begin{aligned}
W_1\big(\mathbb{P}(\cdot|x), \mathbb{P}(\cdot|y)\big) &= W_1\big(\mathbb{P}(\cdot|x_1), \mathbb{P}(\cdot|x_{k+1})\big) \\
&\leq \sum_{j=1}^{k} W_1\big(\mathbb{P}(\cdot|x_j), \mathbb{P}(\cdot|x_{j+1})\big) \\
&\leq (1 - \kappa)k \\
&= (1 - \kappa)\rho_{\mathrm{H}}(x, y),
\end{aligned}$$

where the second step follows from the triangle inequality and the third step follows from (46). Using this observation, we can prove the following:

Lemma 3. *Let $\mathbb{P}_1, \ldots, \mathbb{P}_{T+1}$ be the Markov kernels on X given by (25), and let $\pi_1, \ldots, \pi_{T+1} \in \mathscr{P}(X)$ be the Gibbs measures defined in (19). Suppose that the parameter $\beta > 0$ satisfies $\beta\Delta < 1$. Then, the conditions of Lemma 2 are satisfied with*

$$\delta_t = \frac{\beta |V|^2 (\Delta + 1)}{t + 1} \tag{47}$$

and

$$\kappa^\star = \frac{1 - \Delta\beta}{|V|}. \tag{48}$$

Proof. The fact that each π_t is invariant with respect to $\mathbb{P}_t$ follows from the detailed balance property (26). To keep the paper relatively self-contained, we give in Appendix B a short proof of (26) as a consequence of a more general result on the Gibbs sampler.

To upper-bound the Wasserstein distance $W_1(\pi_t, \pi_{t+1})$, we write

$$\begin{aligned}
W_1(\pi_t, \pi_{t+1}) &= \inf_{v \in C(\pi_t, \pi_{t+1})} \int_{X \times X} \rho_{\mathrm{H}}(x, y) v(\mathrm{d}x, \mathrm{d}y) \\
&\leq |V| \int_{X \times X} \mathbf{1}_{\{x \neq y\}} v(\mathrm{d}x, \mathrm{d}y) \\
&= |V| \cdot \|\pi_t - \pi_{t+1}\|_{\mathrm{TV}}, \tag{49}
\end{aligned}$$

where in the first line we have used the definition (36) of the Wasserstein distance, while the last step follows from the coupling representation (38) of the total variation distance. Furthermore, using the CKKP inequality (8) and the relative-entropy bound (21), we get

$$\|\pi_t - \pi_{t+1}\|_{\mathrm{TV}} \le \sqrt{\frac{1}{2} D(\pi_t \| \pi_{t+1})} \le \frac{\beta |V|(\Delta + 1)}{t+1} \qquad \text{for all } t \ge 1.$$

Using this bound in (49), we get (47).

Finally, we obtain a uniform lower bound on the Ricci curvature of the $\mathbb{P}_t$'s. Each $\mathbb{P}_t$ is of the form

$$\mathbb{P}_t(y|x) = \frac{1}{|V|} \sum_{v \in V} \mathbb{P}_{v,t}(y_v | x_{\partial v}) \mathbf{1}_{\{y_{-v} = x_{-v}\}},$$

where

$$\mathbb{P}_{v,t}(y_v | x_{\partial v}) = \frac{\mu_{v,0}(y_v) \exp\left(-\beta F_{v,t-1}(y_v, x_{\partial v})\right)}{Z_{v,t}(x_{\partial v})}. \tag{50}$$

Recalling the discussion preceding the statement of the lemma, we only need to consider pairs x, y with $\rho_{\mathrm{H}}(x, y) = 1$. Fix such a pair x, y, and let $u \in V$ denote the single vertex at which they differ. We will construct a suitable coupling of $\mathbb{P}_t(\cdot|x)$ and $\mathbb{P}_t(\cdot|y)$. We define a random couple $(\bar{X}, \bar{Y}) \in \mathsf{X} \times \mathsf{X}$ as follows. Select a vertex $v \in V$ uniformly at random. There are three cases to consider:

- If $v = u$, then $x_{-v} = x_{-u} = y_{-u} = y_{-v}$ and a fortiori

$$\mathbb{P}_{v,t}(\cdot|x_{\partial v}) = \mathbb{P}_{v,t}(\cdot|y_{\partial v}).$$

  In this case, we draw a random sample A from $\mathbb{P}_{u,t}(\cdot|x_{\partial u})$ and let $\bar{X}_u = \bar{Y}_u = A$, $\bar{X}_{-u} = x_{-u}$, $\bar{Y}_{-u} = y_{-u}$. Then $\rho_{\mathrm{H}}(\bar{X}, \bar{Y}) = 0$.

- If $v \in \partial u$, then we sample $(\bar{X}_v, \bar{Y}_v)$ from the optimal coupling of $\mathbb{P}_{v,t}(\cdot|x_{\partial v})$ and $\mathbb{P}_{v,t}(\cdot|y_{\partial v})$ (cf. Remark 1 in Section 4.2.1), and let $\bar{X}_{-v} = x_{-v}$, $\bar{Y}_{-v} = y_{-v}$. Then, we have $\bar{X}_v = \bar{Y}_v$ with probability $1 - \|\mathbb{P}_{v,t}(\cdot|x_{\partial v}) - \mathbb{P}_{v,t}(\cdot|y_{\partial v})\|_{\mathrm{TV}}$, in which case $\rho_{\mathrm{H}}(\bar{X}, \bar{Y}) = \rho_{\mathrm{H}}(x, y)$; on the complementary event $\{\bar{X}_v \ne \bar{Y}_v\}$, the Hamming distance $\rho_{\mathrm{H}}(\bar{X}, \bar{Y})$ will increase to 2.

- If $v \notin \partial_+ u$, then $x_{\partial v} = y_{\partial v}$. We sample a random A from $\mathbb{P}_{v,t}(\cdot|x_{\partial v}) = \mathbb{P}_{v,t}(\cdot|y_{\partial v})$ and let $\bar{X}_v = \bar{Y}_v = A$, $\bar{X}_{-v} = x_{-v}$, and $\bar{Y}_{-v} = y_{-v}$. In this case, $\rho_{\mathrm{H}}(\bar{X}, \bar{Y}) = \rho_{\mathrm{H}}(x, y) = 1$.

Let $\bar{v}$ denote the joint probability distribution of $(\bar{X}, \bar{Y})$. It is easy to show that $\bar{X}$ (respectively, $\bar{Y}$) has distribution $\mathbb{P}_t(\cdot|x)$ (respectively, $\mathbb{P}_t(\cdot|y)$). Therefore, $\bar{v}$ is an element of $C\big(\mathbb{P}_t(\cdot|x), \mathbb{P}_t(\cdot|y)\big)$. Moreover,

$$\int_{\mathsf{X} \times \mathsf{X}} \rho_{\mathrm{H}} d\bar{v} = 0 \cdot \Pr(v = u) + 1 \cdot \Pr(v \notin \partial_+ u) + \sum_{v' \in \partial u} \left(1 + \|\mathbb{P}_{v',t}(\cdot|x_{\partial v'}) - \mathbb{P}_{v',t}(\cdot|y_{\partial v'})\|_{\mathrm{TV}}\right) \Pr(v = v')$$

$$\le 1 - \frac{\Delta + 1}{|V|} + \frac{\Delta(1 + \eta)}{|V|}$$

$$= 1 - \frac{1 - \Delta \eta}{|V|},$$

where

$$\eta = \max_{v \in \partial u} \|\mathbb{P}_{v,t}(\cdot|x_{\partial v}) - \mathbb{P}_{v,t}(\cdot|y_{\partial v})\|_{\mathrm{TV}}.$$

It remains to bound η from above. To that end, we note that, for each $v \in V$, both $\mathbb{P}_{v,t}(\cdot|x_{\partial v})$ and $\mathbb{P}_{v,t}(\cdot|y_{\partial v})$ are Gibbs measures, cf. (50). Therefore, using Lemma 1, we can write

$$\left\| \mathbb{P}_{v,t}(\cdot|x_{\partial v}) - \mathbb{P}_{v,t}(\cdot|y_{\partial v}) \right\|_{\mathrm{TV}} \leq \frac{\beta \left\| F_{v,t-1}(\cdot, x_{\partial v}) - F_{v,t-1}(\cdot, y_{\partial v}) \right\|_{\mathrm{s}}}{4}. \tag{51}$$

Using Lemma D.1 in Appendix D and an argument similar to the one used to derive (34), we get

$$\begin{aligned}
\left\| F_{v,t-1}(\cdot, x_{\partial v}) - F_{v,t-1}(\cdot, y_{\partial v}) \right\|_{\mathrm{s}} &\leq 2 \left\| F_{v,t-1}(\cdot, x_{\partial v}) - F_{v,t-1}(\cdot, y_{\partial v}) \right\|_{\infty} \\
&\leq \frac{2}{t} \sum_{s=1}^{t-1} \| f_{v,s}(\cdot, x_{\partial v}) - f_{v,s}(\cdot, y_{\partial v}) \|_{\infty} \\
&\leq 4 \rho_{\mathrm{H}}(x_{\partial v}, y_{\partial v}).
\end{aligned}$$

Since x and y differ only at a single vertex, we have that $\rho_{\mathrm{H}}(x_{\partial v}, y_{\partial v}) \leq 1$. Therefore,

$$\left\| F_{v,t-1}(\cdot, x_{\partial v}) - F_{v,t-1}(\cdot, y_{\partial v}) \right\|_{\mathrm{s}} \leq 4.$$

Note that this bound is *independent* of t; this is a consequence of the $\frac{1}{t+1}$ scaling in Eq. (20), which is in turn a direct consequence of the relative-entropy term in the instantaneous losses. Substituting this into (51), we get

$$\left\| \mathbb{P}_{v,t}(\cdot|x_{\partial v}) - \mathbb{P}_{v,t}(\cdot|y_{\partial v}) \right\|_{\mathrm{TV}} \leq \beta.$$

Therefore, by the definition of W_1 it follows that

$$W_1\left(\mathbb{P}_t(\cdot|x), \mathbb{P}_t(\cdot|y) \right) \leq \int_{\mathsf{X} \times \mathsf{X}} \rho_{\mathrm{H}} \mathrm{d}\bar{v} \leq 1 - \frac{1 - \Delta\beta}{|V|},$$

which in view of relation (46) yields (48). □

4.2.3 Completing the proof

We decompose the regret $R_T^{\mathrm{LI}}(f^T)$ as follows:

$$\begin{aligned}
R_T^{\mathrm{LI}}(f^T) &= \sum_{t=1}^{T} \ell_t(\mu_t) - \inf_{v \in \mathscr{P}(\mathsf{X})} \sum_{t=1}^{T} \ell_t(v) \\
&= \sum_{t=1}^{T} \left(\ell_t(\mu_t) - \ell_t(\pi_t) \right) + \sum_{t=1}^{T} \ell_t(\pi_t) - \inf_{v \in \mathscr{P}(\mathsf{X})} \sum_{t=1}^{T} \ell_t(v) \\
&\leq \sum_{t=1}^{T} \left(\ell_t(\mu_t) - \ell_t(\pi_t) \right) + R_T(f^T). \tag{52}
\end{aligned}$$

Next, we use the form of the instantaneous costs ℓ_t to expand the first summation on the right-hand side of (52):

$$\sum_{t=1}^{T} \left(\ell_t(\mu_t) - \ell_t(\pi_t) \right) = \sum_{t=1}^{T} \beta \left(\langle \mu_t, f_t \rangle - \langle \pi_t, f_t \rangle \right) + \sum_{t=1}^{T} \left(D(\mu_t \| \mu_0) - D(\pi_t \| \mu_0) \right). \tag{53}$$

By Lemma D.2, each f_t is Lipschitz with respect to the Hamming metric with constant $2|V|(\Delta+1)$. Therefore, using Corollary 1, we obtain

$$\langle \mu_t, f_t \rangle - \langle \pi_t, f_t \rangle \le \|f\|_{\text{Lip}}\left((1-\kappa^\star)^{t-1}W_1(\mu_1,\pi_1) + \mathbf{1}_{\{t\ge2\}}\sum_{s=1}^{t-1}(1-\kappa^\star)^{t-1-s}\delta_s\right)$$

$$\le 2|V|(\Delta+1)\left((1-\kappa^\star)^{t-1}W_1(\mu_1,\pi_1) + \mathbf{1}_{\{t\ge2\}}\sum_{s=1}^{t-1}(1-\kappa^\star)^{t-1-s}\delta_s\right).$$

Using the expression for δ_t given in Lemma 3, we further obtain

$$\langle \mu_t, f_t \rangle - \langle \pi_t, f_t \rangle \le 2|V|(\Delta+1)\left((1-\kappa^\star)^{t-1}W_1(\mu_1,\pi_1) + \mathbf{1}_{\{t\ge2\}}\sum_{s=1}^{t-1}(1-\kappa^\star)^{t-1-s}\beta|V|^2(\Delta+1)\gamma_{s+1}\right).$$

Therefore,

$$\sum_{t=1}^{T}\beta\left(\langle \mu_t, f_t \rangle - \langle \pi_t, f_t \rangle\right) \le 2\beta|V|(\Delta+1)W_1(\mu_1,\pi_1)\sum_{t=1}^{T}(1-\kappa^\star)^{t-1} + 2\beta^2|V|^3(\Delta+1)^2\sum_{t=2}^{T}\sum_{s=1}^{t-1}(1-\kappa^\star)^{t-1-s}\gamma_{s+1}$$

$$\le 2\beta|V|(\Delta+1)W_1(\mu_1,\pi_1)\frac{1}{\kappa^\star} + 2\beta^2|V|^3(\Delta+1)^2\sum_{\tau=1}^{T-1}\sum_{s=1}^{\tau}(1-\kappa^\star)^{\tau-s}\gamma_{s+1},$$

$$(54)$$

where the last inequality is obtained by using

$$\sum_{t=1}^{T}(1-\kappa^\star)^{t-1} \le \sum_{t=1}^{\infty}(1-\kappa^\star)^{t-1} = \frac{1}{k^\star}, \tag{55}$$

and by letting $\tau = t-1$ in the second sum over t. By exchanging the order of summation in the last term in (54) and using (55), we have

$$\sum_{\tau=1}^{T-1}\sum_{s=1}^{\tau}(1-\kappa^\star)^{\tau-s}\gamma_{s+1} = \sum_{t=2}^{T}\gamma_t\sum_{\tau=0}^{T-t}(1-\kappa^\star)^{\tau} \le \frac{1}{\kappa^\star}\sum_{t=2}^{T}\gamma_t,$$

implying that

$$\sum_{t=1}^{T}\beta\left(\langle \mu_t, f_t \rangle - \langle \pi_t, f_t \rangle\right) \le 2\beta|V|(\Delta+1)W_1(\mu_1,\pi_1)\frac{1}{\kappa^\star} + 2\beta^2|V|^3(\Delta+1)^2\frac{1}{\kappa^\star}\sum_{t=2}^{T}\gamma_t.$$

Since $\gamma_t = \frac{1}{t}$ for all $t \ge 1$, it follows that

$$\sum_{t=1}^{T}\beta\left(\langle \mu_t, f_t \rangle - \langle \pi_t, f_t \rangle\right) \le 2\beta|V|(\Delta+1)W_1(\mu_1,\pi_1)\frac{1}{\kappa^\star} + 2\beta^2|V|^3(\Delta+1)^2\frac{1}{\kappa^\star}\sum_{t=2}^{T}\frac{1}{t}$$

$$\le 2\beta|V|(\Delta+1)W_1(\mu_1,\pi_1)\frac{1}{\kappa^\star} + 2\beta^2|V|^3(\Delta+1)^2\frac{1}{\kappa^\star}\ln T, \tag{56}$$

where the last inequality follows from

$$\sum_{t=2}^{T}\frac{1}{t} \le \int_1^T\frac{\mathrm{d}t}{t} = \ln T.$$

23

Next, we deal with the relative entropy difference term in (53). Given a probability distribution $\mu \in \mathcal{P}(\mathsf{X})$, let $H(\mu) = -\langle \mu, \ln \mu \rangle$ denote its *Shannon entropy* [10]. Then

$$D(\mu_t \| \mu_0) - D(\pi_t \| \mu_0) = H(\pi_t) - H(\mu_t) + \left\langle \mu_t, \ln \frac{1}{\mu_0} \right\rangle - \left\langle \pi_t, \ln \frac{1}{\mu_0} \right\rangle$$

$$\leq |H(\pi_t) - H(\mu_t)| + \|\pi_t - \mu_t\|_{\mathrm{TV}} \cdot |V| \ln \frac{1}{\theta_d}, \tag{57}$$

where $\theta_d = \min_{v \in V} \min_{a \in \{1,\dots,q\}} \mu_{v,0}(a)$. To upper-bound the first term in (57), we use the following continuity estimate for the Shannon entropy (see, e.g., [10, Theorem 17.3.3]): For any two $\mu, \nu \in \mathcal{P}(\mathsf{X})$ with $\|\mu - \nu\|_{\mathrm{TV}} \leq 1/4$,

$$|H(\mu) - H(\nu)| \leq 2\left(\|\mu - \nu\|_{\mathrm{TV}} \ln |\mathsf{X}| + \|\mu - \nu\|_{\mathrm{TV}} \ln \frac{1}{\|\mu - \nu\|_{\mathrm{TV}}} \right),$$

where $|\mathsf{X}| = q^{|V|}$ is the cardinality of X. In order to use this estimate, we need an upper bound on $\|\pi_t - \mu_t\|_{\mathrm{TV}}$, which can be obtained as follows:

$$\|\pi_t - \mu_t\|_{\mathrm{TV}} = \inf_{\upsilon \in C(\pi_t, \mu_t)} \int_{\mathsf{X} \times \mathsf{X}} \mathbf{1}_{\{x \neq y\}} \upsilon(\mathrm{d}x, \mathrm{d}y)$$

$$\leq \inf_{\upsilon \in C(\pi_t, \mu_t)} \int_{\mathsf{X} \times \mathsf{X}} \rho_{\mathrm{H}}(x, y) \upsilon(\mathrm{d}x, \mathrm{d}y)$$

$$= W_1(\pi_t, \mu_t)$$

$$\leq (1 - \kappa^\star)^{t-1} W_1(\mu_1, \pi_1) + \mathbf{1}_{\{t \geq 2\}} \sum_{s=1}^{t-1} (1 - \kappa^\star)^{t-1-s} \delta_s, \tag{58}$$

where the last step follows from Lemma 2 (see (42)). We can upper-bound the Wasserstein distance $W_1(\mu_1, \pi_1)$ as follows:

$$W_1(\mu_1, \pi_1) = \inf_{\upsilon \in C(\mu_1, \pi_1)} \int_{\mathsf{X}} \rho_{\mathrm{H}}(x, y) \upsilon(\mathrm{d}x, \mathrm{d}y)$$

$$\leq |V| \inf_{\upsilon \in C(\mu_1, \pi_1)} \int_{\mathsf{X}} \mathbf{1}_{\{x \neq y\}} \upsilon(\mathrm{d}x, \mathrm{d}y)$$

$$= |V| \|\mu_1 - \pi_1\|_{\mathrm{TV}}$$

$$\leq |V|.$$

Using this and the fact that $\delta_s \leq \frac{\beta |V|^2 (\Delta+1)}{s+1}$ by Lemma 3 in (58), we can write

$$\|\pi_t - \mu_t\|_{\mathrm{TV}} \leq K \sum_{s=1}^{t} \frac{(1 - \kappa^\star)^{t-s}}{s}$$

$$= K p_t (1 - \kappa^\star), \tag{59}$$

where $K \equiv K(\beta, |V|, \Delta) \triangleq \max\{|V|, \beta |V|^2 (\Delta + 1)\}$, and $p_t(u) \triangleq \sum_{s=1}^{t} \frac{u^{t-s}}{s}$. As a consequence of Lemma C.1 in Appendix C, there exists a finite $T_0 = T_0(\beta, |V|, \Delta + 1)$, such that the sequence $\{p_t(1 - \kappa^\star)\}_{t=T_0}^{\infty}$ is strictly decreasing and convergent to zero. Therefore, there exists a finite $T_1 \equiv T_1(\beta, |V|, \Delta)$, such that

$$\|\pi_t - \mu_t\|_{\mathrm{TV}} \leq K p_t (1 - \kappa^\star) \leq \frac{1}{4}, \qquad t \geq T_1.$$

Moreover, the function $u \mapsto -u \ln u$ is increasing on the open interval $(0, 1/e)$, so for $t \geq T_1$ we have

$$\|\pi_t - \mu_t\|_{\mathrm{TV}} \ln \frac{1}{\|\mu_t - \pi_t\|_{\mathrm{TV}}} \leq K p_t (1 - \kappa^\star) \ln \frac{1}{K p_t (1 - \kappa^\star)}$$
$$\leq K p_t (1 - \kappa^\star) \ln t,$$

where we have also used the fact that $p_t(u) \geq 1/t$. Consequently,

$$D(\mu_t \| \pi_0) - D(\pi_t \| \pi_0) \leq \begin{cases} |V| \ln \frac{q^2}{\theta_d}, & t < T_1 \\ K p_t (1 - \kappa^\star) \left(|V| \ln \frac{q^2}{\theta_d} + \ln t \right), & t \geq T_1 \end{cases}. \tag{60}$$

Summing from $t = 1$ to $t = T$, we get

$$\sum_{t=1}^{T} \left[D(\mu_t \| \pi_0) - D(\pi_t \| \pi_0) \right] \leq T_1 |V| \ln \frac{q^2}{\theta_d} + K \left(|V| \ln \frac{q^2}{\theta_d} + \ln T \right) \sum_{t=1}^{T} p_t (1 - \kappa^\star)$$

$$= T_1 |V| \ln \frac{q^2}{\theta_d} + K \left(|V| \ln \frac{q^2}{\theta_d} + \ln T \right) \sum_{t=1}^{T} \sum_{s=1}^{t} \frac{(1 - \kappa^\star)^{t-s}}{s}$$

$$\leq T_1 |V| \ln \frac{q^2}{\theta_d} + K \left(|V| \ln \frac{q^2}{\theta_d} + \ln T \right) \frac{1}{\kappa^\star} \ln(T + 1). \tag{61}$$

Combining (56) and (61), we get

$$\sum_{t=1}^{T} \left[\ell_t(\mu_t) - \ell_t(\pi_t) \right]$$

$$\leq \sum_{t=1}^{T} \left[\beta \left| \langle \mu_t, f_t \rangle - \langle \pi_t, f_t \rangle \right| + \left| D(\mu_t \| \mu_0) - D(\pi_t \| \mu_0) \right| \right]$$

$$\leq \frac{2\beta |V|^2 (\Delta + 1)}{\kappa^\star} + T_1 |V| \ln \frac{q^2}{\theta_d} + \frac{1}{\kappa^\star} \left(2\beta^2 |V|^3 (\Delta + 1)^2 + K \left(|V| \ln \frac{q^2}{\theta_d} + \ln T \right) \right) \ln(T + 1)$$

We can now obtain the bound on the overall regret, via (52):

$$R_T^{\mathrm{LI}}(f^T) \leq \frac{1}{\kappa^\star} \left(2\beta^2 |V|^3 (\Delta + 1)^2 + K \left(|V| \ln \frac{q^2}{\theta_d} + \ln T \right) \right) \ln(T + 1)$$

$$+ 2 \left(\beta |V| (\Delta + 1) \right)^2 \ln(T + 1) + T_1 |V| \ln \frac{q^2}{\theta_d} + \frac{2\beta |V|^2 (\Delta + 1)}{\kappa^\star} + \ln \frac{1}{\theta}. \tag{62}$$

Substituting the expression for $\kappa^\star$ from Lemma 3, we get (28), and the proof is complete.

5 Conclusion

We have studied a model of online (i.e., real-time) discrete optimization by a social network consisting of agents that must choose actions to balance their immediate time-varying costs against a tendency to act according to some default myopic strategy. The costs are generated by a dynamic environment, and the agents lack ability or incentive to construct an a priori model of the environment's evolution. The global cost of the network decomposes into a sum of individual and pairwise local-interaction terms and, at each time step, every agent is informed only about its own cost and the pairwise costs in its

immediate neighborhood. These assumptions on the network and on the environment capture the so-called *Knightian uncertainty* [7–9]. The overall objective is to minimize the worst-case regret, i.e., the difference between the cumulative real-time performance of the network and the best performance that could have been achieved in hindsight with full centralized knowledge. We have constructed an explicit strategy for the network based on the Glauber dynamics and showed that it achieves favorable scaling of the regret in terms of problem parameters under a Dobrushin-type mixing condition. Our proof uses ideas from statistical physics, as well as recent developments in the theory of Markov chains in metric spaces, specifically Ollivier's notion of positive Ricci curvature of a Markov operator [28].

Although the notion of regret is backward-looking, it is important conceptually since it quantifies the agents' ability to make *forecasts* even in the absence of a Bayesian model, and to improve their decisions over time. From the point of view of economics, regret minimization is significant for two reasons. First, from the positive (or descriptive) standpoint, it allows for boundedly rational agents. Second, it may be used as a basis for what Selten [54] has called a *practically normative* theory of economic behavior, since the goal of minimizing regret is synonymous with using past experience to improve one's decisions in the future, as opposed to following a strategy based on ideal rational expectations independent of the environment. In addition, in the online learning framework, the model of the interaction between the social network and the environment does not rely on probability judgments or assumptions about *what will happen*. Rather, probability is used as a tool to help the agents decide *what to do* – how to allocate priority to different actions? When to perform experimentation, and when to stick with a strategy that had performed well in the past? Thus, probability is used as an objective *evolutionary mechanism* for selecting an action [54, 55] or as a mechanism to keep track of past experience in a case-based decision framework [56], rather than as a subjective *belief* about the environment. This viewpoint is, of course, ideally suited for a Knightian theory of decision-making, and it meshes well with post-Keynesian critiques of the use of probability to quantify uncertainty [57, 58].

A   Proof of Lemma 1

All Gibbs measures μ_g induced by the same base measure μ have the same support as μ. Therefore, the quantity $D(\mu_g \| \mu_h)$ is finite for all functions g and h on X, and

$$
\begin{aligned}
D(\mu_g \| \mu_h) &= \left\langle \mu_g, \ln \frac{\mu_g}{\mu_h} \right\rangle \\
&= \langle \mu_g, g - h \rangle + \ln \frac{\langle \mu, \exp(h) \rangle}{\langle \mu, \exp(g) \rangle} \\
&= \langle \mu_g, g - h \rangle + \ln \langle \mu_g, \exp(h - g) \rangle .
\end{aligned} \tag{A.1}
$$

We now use the well-known Hoeffding bound [59], which for our purposes can be stated as follows: for any function $F : \mathsf{X} \to \mathbb{R}$ and any $v \in \mathscr{P}(\mathsf{X})$,

$$
\ln \langle v, \exp(F) \rangle \leq \langle v, F \rangle + \frac{\|F\|_{\mathrm{s}}^2}{8} . \tag{A.2}
$$

Applying (A.2) to the second term in (A.1), we note that the terms involving the expectation of $g - h$ with respect to μ_g cancel, and we are left with (10). The bound (11) follows from (10) and the CKKP inequality (8).

B Gibbs sampler and detailed balance

In order to keep the paper self-contained, we give a brief proof of the detailed balance property of the discrete-state Gibbs sampler [23]. Consider an arbitrary everywhere positive probability measure $\pi \in \mathscr{P}(\mathsf{X})$ and a random variable $X = (X_v)_{v \in V}$ with distribution π. For any $v \in V$, the conditional probability that $X_v = x_v$ given $X_{-v} = x_{-v}$ is equal to

$$\pi_v(x_v | x_{-v}) \triangleq \frac{\pi(x_v, x_{-v})}{\pi_{-v}(x_{-v})},$$

where (a, x_{-v}) denotes the tuple $y \in \mathsf{X}$ obtained from x by replacing x_v with a, i.e., $y_v = a$ and $y_{-v} = x_{-v}$, and

$$\pi_{-v}(x_{-v}) = \sum_{a \in \{1, \dots, q\}} \pi(a, x_{-v}).$$

The *Gibbs sampler* is implemented as follows: starting from $x \in \mathsf{X}$, pick a vertex $v \in V$ uniformly at random, replace x_v with a random sample Y_v from $\pi_v(\cdot | x_{-v})$, and let $Y_{-v} = x_{-v}$. The overall stochastic transformation $x \to Y$ is described by the Markov kernel

$$\mathbb{P}(y | x) = \frac{1}{|V|} \sum_{v \in V} \pi_v(y_v | x_{-v}) \mathbf{1}_{\{x_{-v} = y_{-v}\}}.$$

Then we claim that the pair $(\pi, \mathbb{P})$ has the detailed balance property

$$\pi(x)\mathbb{P}(y | x) = \pi(y)\mathbb{P}(x | y), \qquad \forall x, y \in \mathsf{X}.$$

Indeed,

$$\begin{aligned}
\pi(x)\mathbb{P}(y | x) &= \frac{1}{|V|} \sum_{v \in V} \pi_v(y_v | x_{-v}) \pi(x_v, y_{-v}) \mathbf{1}_{\{x_{-v} = y_{-v}\}} \\
&= \frac{1}{|V|} \sum_{v \in V} \frac{\pi(y_v, x_{-v})}{\pi_{-v}(x_{-v})} \pi(x_v, y_{-v}) \mathbf{1}_{\{x_{-v} = y_{-v}\}} \\
&= \frac{1}{|V|} \sum_{v \in V} \frac{\pi(y_v, x_{-v})}{\pi_{-v}(y_{-v})} \pi(x_v, y_{-v}) \mathbf{1}_{\{x_{-v} = y_{-v}\}} \\
&= \frac{1}{|V|} \sum_{v \in V} \frac{\pi(x_v, y_{-v})}{\pi_{-v}(y_{-v})} \pi(y_v, x_{-v}) \mathbf{1}_{\{x_{-v} = y_{-v}\}} \\
&= \frac{1}{|V|} \sum_{v \in V} \pi_{-v}(x_v | y_{-v}) \pi(y_v, x_{-v}) \mathbf{1}_{\{x_{-v} = y_{-v}\}} \\
&= \pi(y)\mathbb{P}(x | y).
\end{aligned}$$

A simple calculation shows that when $\pi = \mu_f$ for a Gibbs measure μ_f induced by an everywhere positive product measure $\mu \in \mathscr{P}(\mathsf{X})$ and any function $f \in \mathscr{F}$, the conditional measure $\pi_v(\cdot | x_{-v})$ for any $v \in V$ has the form

$$\pi_{-v}(\cdot | x_{-v}) \propto \mu_v(x_v) \exp\left(-f_v(\cdot, x_{\partial v})\right).$$

This, in turn, implies the detailed balance property (26).

C A polynomial recurrence in the proof of Theorem 1

For each $t = 1, 2, \ldots$, consider the polynomial

$$p_t(u) = \sum_{s=1}^{t} \frac{u^{t-s}}{s}.$$

We are interested in its behavior on the interval $[0, 1]$.

Lemma C.1. *For each $u \in [0, 1)$, the sequence $\{p_t(u)\}_{t=1}^{\infty}$ converges to zero. Moreover, there exists a finite $t_0 = t_0(u) \in \mathbb{N}$, such that $p_{t+1}(u) < p_t(u)$ for all $t \geq t_0$.*

Proof. We first observe the following recurrence relation: for any $u \in [0, 1]$,

$$p_{t+1}(u) = u p_t(u) + \frac{1}{t+1}. \tag{C.1}$$

From this, we see that $\lim_{t \to \infty} p_t(u) = 0$ for any $u \in [0, 1)$. Let us fix some such u. Suppose that $p_{t_0+1}(u) < p_{t_0}(u)$ for some t_0. Then we claim that $p_{s+1}(u) < p_s(u)$ for all $s \geq t_0$. Indeed, from (C.1),

$$
\begin{aligned}
p_{t_0+2}(u) &= u p_{t_0+1}(u) + \frac{1}{t_0 + 2} \\
&< u p_{t_0}(u) + \frac{1}{t_0 + 2} \\
&< u p_{t_0}(u) + \frac{1}{t_0 + 1} \\
&= p_{t_0+1}(u).
\end{aligned}
$$

The general claim of strict monotonicity then follows by induction. It remains to prove that such a finite t_0 always exists. To that end, consider for arbitrary t the polynomial

$$q_t(u) \triangleq p_{t+1}(u) - p_t(u) = u^{t+1} - \sum_{s=1}^{t-1} \frac{u^{t+1-s}}{s(s+1)}.$$

The leading coefficient of q_t is positive while all other coefficients are negative, so, by Descartes' rule of signs, q_t has exactly one positive real root. Let us denote this root by u_t. We claim that $u_t \in (0, 1]$. Indeed, u_t must be positive, since $q_t(u)$ has a nonzero constant term. Moreover, $q_t(0) = -\frac{1}{t(t+1)}$ and $q_t(1) = \frac{1}{t+1}$, so q_t changes sign in $[0, 1]$. Thus, $u_t \in (0, 1]$, and

$$p_{t+1}(u) < p_t(u), \qquad u < u_t.$$

By virtue of this strict monotonicity property, the sequence $\{u_t\}_{t=1}^{\infty}$ is strictly increasing and bounded by one. Now, for a given u simply take t_0 to be the smallest element of the set $\{t \in \mathbb{N} : u_t > u\}$. $\qquad\square$

D Miscellanea

Lemma D.1. *Consider all functions $f : X \to \mathbb{R}$ of the form* (13), *where all local terms ϕ_v and $\phi_{u,v}$ take values in the interval $[-1, 1]$. Then*

$$\|f\|_{\infty} \leq |V|(\Delta + 1), \tag{D.1}$$

where $\|f\|_\infty \triangleq \max_{x\in\mathsf{X}}|f(x)|$ is the sup norm of f. Moreover, for any $x, y \in \mathsf{X}$ and any $v \in V$,

$$\left\|f_v(\cdot, x_{\partial v}) - f_v(\cdot, y_{\partial v})\right\|_\infty \le 2\rho_{\mathrm{H}}(x_{\partial v}, y_{\partial v}), \tag{D.2}$$

where

$$\rho_{\mathrm{H}}(x_{\partial v}, y_{\partial v}) = \sum_{u\in\partial v} \mathbf{1}_{\{x_u \ne y_u\}}.$$

Proof. For any $x \in \mathsf{X}$, we have

$$|f(x)| \le \sum_{v\in V} \left|\phi_v(x_v)\right| + \sum_{\{u,v\}\in E} \left|\psi_{u,v}(x_u, x_v)\right|$$

$$\le |V| + |E|.$$

Since the graph $G = (V, E)$ is undirected and simple, an elementary counting argument shows that

$$|E| \le |V|\Delta/2.$$

Overbounding slightly, we get (D.1). Similarly, for any $a \in \{1, \dots, q\}$,

$$\left|f_v(a, x_{\partial v}) - f_v(a, y_{\partial v})\right| \le \sum_{u\in\partial v} \left|\psi_{u,v}(a, x_{\partial v}) - \psi_{u,v}(a, y_{\partial v})\right|$$

$$\le 2 \sum_{u\in\partial v} \mathbf{1}_{\{x_u \ne y_u\}}$$

$$= 2\rho_{\mathrm{H}}(x_{\partial v}, y_{\partial v}),$$

which gives us (D.2). $\qquad\square$

Lemma D.2. *Under the same assumptions as in Lemma D.1, each cost function f of the form (13) is Lipschitz with respect to the Hamming distance ρ_{H}, with Lipschitz constant $\|f\|_{\mathrm{Lip}} \le 2|V|(\Delta + 1)$.*

Proof. For any two $x, y \in \mathsf{X}$, we have

$$\left|f(x) - f(y)\right| \le \sum_{v\in V} \left|\phi_v(x_v) - \phi_v(y_v)\right| + \sum_{\{u,v\}\in E} \left|\psi_{u,v}(x_u, x_v) - \psi_{u,v}(y_u, y_v)\right|$$

$$\le 2\left\{\sum_{v\in V} \mathbf{1}_{\{x_v \ne y_v\}} + \sum_{\{u,v\}\in E} \mathbf{1}_{\{(x_u, x_v) \ne (y_u, y_v)\}}\right\}$$

$$\le 2\left\{\sum_{v\in V} \mathbf{1}_{\{x_v \ne y_v\}} + \sum_{\{u,v\}\in E} \left(\mathbf{1}_{\{x_u \ne y_u\}} + \mathbf{1}_{\{x_v \ne y_v\}}\right)\right\}$$

$$\le 2(|V| + 2|E|) \sum_{v\in V} \mathbf{1}_{\{x_v \ne y_v\}}$$

$$= 2|V|(\Delta + 1)\rho_{\mathrm{H}}(x, y).$$

$\qquad\square$

References

[1] C. P. Chamley. *Rational Herds: Economic Models of Social Learning*. Cambridge Univ. Press, 2004.

[2] J. C. Harsanyi. Games with incomplete information played by "Bayesian" players. *Management Science*, 14(3):159–182, 1967.

[3] E. Kalai and E. Lehrer. Rational learning leads to Nash equilibrium. *Econometrica*, 61(5):1019–1045, 1993.

[4] M. O. Jackson. *Social and Economic Networks*. Princeton Univ. Press, 2010.

[5] D. Acemoglu, M. A. Dahleh, I. Lobel, and A. Ozdaglar. Bayesian learning in social networks. *Review of Economic Studies*, 78:1201–1236, 2011.

[6] A. Jadbabaie, P. Molavi, A. Sandroni, and A. Tahbaz-Salehi. Non-Bayesian social learning. *Games and Economic Behavior*, 76:210–225, 2012.

[7] F. H. Knight. *Risk, Uncertainty and Profit*. Houghton Mifflin, Boston, 1921.

[8] T. F. Bewley. Knightian decision theory. Part I. *Decisions in Economics and Finance*, 25:79–110, 2002.

[9] T. F. Bewley. Knightian decision theory. Part II: Intertemporal problems. Cowles Foundation for Research in Economics discussion paper 835, Yale University, May 1987.

[10] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, New York, 2nd edition, 2006.

[11] D. McFadden. Conditional logit analysis of qualitative choice behavior. In P. Zarembka, editor, *Frontiers in Econometrics*, pages 105–142. Academic Press, 1974.

[12] L. P. Hansen and T. J. Sargent. *Robustness*. Princeton University Press, 2008.

[13] J. Abernethy, A. Agarwal, P. L. Bartlett, and A. Rakhlin. A stochastic view of optimal regret through minimax duality. In *Proc. Conf. on Learning Theory*, 2009.

[14] D. P. Foster and R. Vohra. Regret in the on-line decision problem. *Games and Economic Behavior*, 29:7–35, 1999.

[15] N. Cesa-Bianchi and G. Lugosi. *Prediction, Learning and Games*. Cambridge Univ. Press, 2006.

[16] J. Abernethy, R. M. Froniglio, and A. Wibisono. Minimax option pricing meets Black–Scholes in the limit. In *Proc. 44th Symposium on Theory of Computing (STOC)*, pages 1029–1040, 2012.

[17] J. Abernethy, Y. Chen, and J. Wortman Vaughan. Efficient market making via convex optimization, and a connection to online learning. *ACM Transactions on Economics and Computation*, 1(2):12:1–12:39, May 2013.

[18] J. Hannan. Approximation to Bayes risk in repeated play. In M. Dresher, A. W. Tucker, and P. Wolfe, editors, *Contributions to the Theory of Games*, volume 3, pages 97–139. Princeton University Press, 1957.

[19] E. Takimoto and M. K. Warmuth. Path kernels and multiplicative updates. *J. Machine Learning Res.*, 4:773–818, 2003.

[20] A. Kalai and S. Vempala. Efficient algorithms for online decision problems. *J. Comp. Sys. Sci.*, 71:291–307, 2005.

[21] R. Glauber. Time-dependent statistics of the Ising model. *J. Math. Phys.*, 4:294–307, 1963.

[22] V. F. Turchin. On the computation of multidimensional integrals by the Monte Carlo method. *Theory Probab. Appl.*, 16(4):720–724, 1971.

[23] L. Tierney. Markov chains for exploring posterior distributions. *Ann. Statist.*, 22(4):1701–1762, 1994.

[24] L. E. Blume. The statistical mechanics of strategic interaction. *Games and Economic Behavior*, 5:387–424, 1993.

[25] H. Peyton Young. *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions.* Princeton Univ. Press, 2001.

[26] C. Alós-Ferrer and N. Netzer. The logit-response dynamics. *Games and Economic Behavior*, 68:413–427, 2010.

[27] W. H. Sandholm. *Population Games and Evolutionary Dynamics.* MIT Press, 2010.

[28] Y. Ollivier. Ricci curvature of Markov chains on metric spaces. *J. Funct. Anal.*, 256:810–864, 2009.

[29] D. Gamarnik, D. A. Goldberg, and T. Weber. Correlation decay in random decision networks. *Math. Oper. Res.*, 39(2):229–261, 2014.

[30] B. Simon. *The Statistical Mechanics of Lattice Gases*, volume 1. Princeton University Press, Princeton, NJ, 1993.

[31] D. Foster and R. Vohra. Calibrated learning and correlated equilibrium. *Games and Economic Behavior*, 21:40–55, 1997.

[32] D. Foster and H. Peyton Young. Learning, hypothesis testing, and Nash equilibrium. *Games and Economic Behavior*, 45:73–96, 2003.

[33] S. Hart and A. Mas-Colell. A simple adaptive procedure leading to correlated equilibrium. *Econometrica*, 68:1127–1150, 2000.

[34] S. Hart and A. Mas-Colell. A general class of adaptive strategies. *Journal of Economic Theory*, 98:26–54, 2001.

[35] M. Kearns. Graphical games. In N. Nisan, T. Roughgarden, É. Tardos, and V. V. Vazirani, editors, *Algorithmic Game Theory*. Cambridge Univ. Press, 2007.

[36] J. Marschak and R. Radner. *Economic Theory of Teams.* Yale University Press, 1972.

[37] R. Radner. Team decision problems. *The Annals of Mathematical Statistics*, 33:857–881, 1962.

[38] M. S. Pinsker. *Information and Information Stability of Random Variables and Processes.* Holden-Day, San Francisco, CA, 1964.

[39] I. Csiszár. Information-type measures of difference of probability distributions and indirect observations. *Stud. Sci. Math. Hung.*, 2:299–318, 1967.

[40] J. H. B. Kemperman. On the optimum rate of transmitting information. *Ann. Math. Statist.*, 40(6):2156–2177, 1969.

[41] S. Kullback. A lower bound for discrimination information in terms of variation. *IEEE Trans. Inform. Theory*, 13:126–217, January 1967.

[42] S. Kullback. Correction to a lower bound for discrimination information in terms of variation. *IEEE Trans. Inform. Theory*, 16:652, September 1970.

[43] P. Bartlett, E. Hazan, and A. Rakhlin. Adaptive online gradient descent. In J. C. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Adv. Neural Inform. Processing Systems*, volume 20, pages 65–72, Cambridge, MA, 2008. MIT Press.

[44] J. Abernethy, P. L. Bartlett, A. Rakhlin, and A. Tewari. Optimal strategies and minimax lower bounds for online convex games. In *Proc. Int. Conf. on Learning Theory*, pages 415–423, 2008.

[45] D. Weitz. Combinatorial criteria for uniqueness of Gibbs measures. *Random Struct. Alg.*, 27(4):445–475, 2005.

[46] M. Dyer, L. A. Goldberg, and M. Jerrum. Matrix norms and rapid mixing for spin systems. *Ann. Appl. Probab.*, 19(1):71–107, 2009.

[47] H. Narayanan and A. Rakhlin. Random walk approach to regret minimization. In J. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. S. Zemel, and A. Culotta, editors, *Advances in Neural Information Processing Systems 23*, pages 1777–1785, 2010.

[48] Yu. E. Nesterov and M. J. Todd. On the Riemannian geometry defined by self-concordant barriers and interior-point methods. *Found. Comput. Math.*, 2:333–361, 2002.

[49] L. Lovász and M. Simonovits. Random walks on a convex body and an improved volume algorithm. *Random Struct. Alg.*, 4(4):359–412, 1993.

[50] A. Joulin and Y. Ollivier. Curvature, concentration and error estimates for Markov chain Monte Carlo. *Ann. Probab.*, 38(6):2418–2442, 2010.

[51] C. Villani. *Topics in Optimal Transportation*, volume 58 of *Graduate Studies in Mathematics*. Amer. Math. Soc., Providence, RI, 2003.

[52] D. A. Levin, Y. Peres, and E. L. Wilmer. *Markov Chains and Mixing Times*. Amer. Math. Soc., 2008.

[53] R. Bubley and M. Dyer. Path coupling: a technique for proving rapid mixing in Markov chains. In *Proc. 38th IEEE Symp. on Foundations of Comp. Sci.*, pages 223–231, 1997.

[54] R. Selten. Evolution, learning, and economic behavior. *Games and Economic Behavior*, 3:3–24, 1991.

[55] R. Nelson and S. G. Winter. *An Evolutionary Theory of Economic Change*. Harvard University Press, 1982.

[56] I. Gilboa and D. Schmeidler. *A Theory of Case-Based Decisions*. Cambridge Univ. Press, 2001.

[57] P. Davidson. Is probability theory relevant for uncertainty? A post Keynesian perspective. *The Journal of Economic Perspectives*, 5(1):129–143, 1991.

[58] J. R. Crotty. Are Keynesian uncertainty and macrotheory incompatible? Conventional decision making, institutional structures and conditional stability in Keynesian macromodels. In G. Dymski and R. Pollin, editors, *New Perspectives in Monetary Macroeconomics: Explorations in the Tradition of Hyman Minsky*, pages 105–142. University of Michigan Press, 1994.

[59] W. Hoeffding. Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Soc.*, 58:13–30, 1963.