Load Balancing Without Regret in the Bulletin Board Model

Robert Kleinberg[*]
Cornell University
rdk@cs.cornell.edu

Georgios Piliouras[†]
Cornell University
gpil@cs.cornell.edu

Éva Tardos[‡]
Cornell University
eva@cs.cornell.edu

ABSTRACT

We analyze the performance of protocols for load balancing in distributed systems based on no-regret algorithms from online learning theory. These protocols treat load balancing as a repeated game and apply algorithms whose average performance over time is guaranteed to match or exceed the average performance of the best strategy in hindsight.

Our approach captures two major aspects of distributed systems. First, in our setting of atomic load balancing, every single process can have a significant impact on the performance and behavior of the system. Furthermore, although in distributed systems participants can query the current state of the system they cannot reliably predict the effect of their actions on it. We address this issue by considering load balancing games in the bulletin board model, where players can find out the delay on all machines, but do not have information on what their experienced delay would have been if they had selected another machine. We show that under these more realistic assumptions, if all players use the well-known multiplicative weights algorithm, then the quality of the resulting solution is exponentially better than the worst correlated equilibrium, and almost as good as that of the worst Nash. These tighter bounds are derived from analyzing the dynamics of a multi-agent learning system.

Categories and Subject Descriptors

F.2.0 and J.4 [**Analysis of Algorithms (General) and Social and Behavior Sciences**]:

[*]Supported by NSF grants CCF-0643934 and CCF-0729102, an Alfred P. Sloan Foundation Fellowship, and a Microsoft Research New Faculty Fellowship.

[†]Supported by NSF grant CCF-0325453.

[‡]Supported by NSF grants CCF-0325453, CCF-0729006, ONR grant N00014-98-1-0589, and a Yahoo! Research Alliance Grant.

PODC'09, August 10–12, 2009, Calgary, Alberta, Canada.

General Terms

Algorithms, Economics, Theory

Keywords

Learning Theory, Game Theory, Price of Anarchy

1. INTRODUCTION

Game theory provides a framework that helps us understand environments where participants interact by selfishly making decisions, and achieve a global outcome without explicit coordination by a single global designer. Modeling various problems from routing, network design, and scheduling as games played by selfish agents has led to many interesting results. Much of this literature uses Nash equilibrium as the solution concept, i.e., defines Nash equilibrium as the outcome in a competitive game. However, Nash equilibrium has several drawbacks that call into question its plausibility as a prediction of a game's outcome. First, the solution concept tells us nothing about the dynamics by which players can be expected to reach an equilibrium. In most games, natural "game play" tends not to converge to Nash equilibria. In fact, the problem of computing Nash equilibria in many games turns out to be computationally hard — it was recently shown to be PPAD-complete [9, 7]. If computing equilibria is computationally hard, it seems unreasonable to assume that players will find such a solution. Further, most games have many equilibria, hence finding one would also involve coordination among the players to agree on one of the possible outcomes.

To overcome these drawbacks, researchers have considered alternative solution concepts based on the long-run average outcome of self-adapting agents who react to each other's strategies in repeated play of the game. Adopting a paradigm known as "sophisticated learning" in the economics literature [13], we assume in this paper that all players use no-regret strategies. No-regret algorithms have the property that in any repeated game play, their average loss per time step approaches that of the single best strategy with hindsight (or better) over time. Regret minimization can be done via simple and efficient strategies, and yet the no-regret property is analogous to the notion of equilibrium (e.g., see the survey of Blum and Mansour [6]). Outcome distributions reached by such no-regret strategies have been well studied, see [3, 13, 26]. If all players play strategies using no-regret algorithms, it results in an empirical distribution of play converging to a weaker set of equilibria: the coarse (weak) correlated equilibria (see below). Foster and

Vohra [16] introduced the stronger notion of *internal regret*, such that no-internal-regret algorithms converge to the set of correlated equilibria.

Our interest in this paper is the quality of outcomes when players play simple regret-minimizing strategies without access to full information about the game. Much of the work in algorithmic game theory concerns the price of anarchy, defined as the ratio of the worst Nash equilibrium to the best outcome, with respect to some global quality measure of the solutions. Blum et al. [4] in PODC'06 were the first to consider the quality of outcome reached when players using no-regret learning strategies. They consider congestion games in the Wardrop (nonatomic) setting of infinitesimal agents, and hence the action of a single player doesn't have significant impact on the system. In this setting they extend the price of anarchy results known for Nash equilibria, to outcomes reached by no-regret learning. Furthermore, Blum at al. [5] defined the price of total anarchy as the ratio of the worst outcome that can be reached by regret minimizing players[1] to the best outcome. Blum et al. [4, 5] show that in some classes of games regret-minimizing players exhibit behavior or global quality that is close to that of a Nash equilibrium. More recently, Roughgarden [24] showed that for a wide class of games (including congestion games that satisfy a natural smoothness condition) bounds on the price of anarchy automatically extend to the total price of anarchy, when the global quality is defined to be the average cost. Unfortunately, in many other classes of games and under natural global quality measures, the price of total anarchy can be significantly worse than the price of anarchy, which suggests that generic no-regret learning is not effective for these games. In fact, the simplest such game is a load balancing game in which there are n jobs and n machines. Each job selects a machine on which to run, and evaluates the outcome as the load of its machine, i.e. the number of jobs on the machine it selected. For this class of games the quality of the worst correlated equilibrium is $\Theta(\sqrt{n})$ [5], whereas a sequence of papers [20, 8, 21] shows that the worst Nash equilibrium is the symmetric fully mixed equilibrium, which has exponentially better quality, namely $\Theta(\log n/\log\log n)$.

In distributed systems we need to consider a further source of difficulty. Traditional learning theory assumes that after playing a round of a game, each player can discover the cost of each possible strategy they could have used given the actions of their opponents. This is a reasonable assumption in games with infinitesimally small players, when actions of a single player have (essentially) no effect on the system. It is also reasonable when the underlying game is well-defined and common knowledge amongst all players. In distributed systems, however, this assumption is rather unnatural. Indeed, different subsystems only need to share some common functionality, whereas their inner workings can vary widely and even be updated seamlessly, and every single process can have significant impact on the behavior of the system.

Despite the pessimistic theoretical predictions mentioned above and the even more demanding setting of applied systems, simple and intuitive adaptive procedures seem to work reasonably well in practice. In this work and in a related paper [19], we analyze models of such network dynamics: we explore the quality of outcomes reached by some concrete learning strategies. We focus on the multiplicative weights learning algorithm (also known as Hedge [12]), which is arguably the simplest and most intuitive no-regret algorithm. In [19] we study the dynamics of these algorithms in atomic congestion games in the traditional full information setting. We show that in almost all such games, the multiplicative-weights learning algorithm results in convergence to pure equilibria. In the game of [19], each process consumes non-negligible system resources and as a result can have a significant impact on the performance and behavior of the whole system. Moving away from the assumption of infinitesimally small users as in [4, 11] adds a significant layer of complexity to the analysis of the system. How much closer can we get to realistic models of applied systems before we become overwhelmed by the complexity of the emerging dynamics [25]?

In this paper, we take a significant additional step towards modeling distributed systems, by moving away from the standard full information setting. We consider load balancing in the so-called "bulletin board model" (similar to the ones in [2, 23]). In this model players can find out the delay on all machines, but do not have information on what their experienced delay would have been if they had selected another machine. Namely, players can query the current state of the system but cannot reliably predict the effect of their actions on it.

Although this change in the players' information might appear benign at first glance, it significantly alters the system behavior. Most importantly, the system becomes symmetric because all players observe the same feedback signal and respond to it using identical algorithms. Thus, at any point in time the players will sample their strategies from identical distributions, and our analysis only needs to focus on how this distribution evolves over time. This is quite different from our analysis of the full information setting in [19], which focused on the *symmetry-breaking* that inevitably occurs when atomic players use the Hedge algorithm in that setting. The symmetric setting that we study here allows for a significantly simpler analysis, incorporating techniques that are standard in the analysis of multiplicative-weight algorithms in learning theory (such as the use of KL-divergence as a potential function) as well as some new techniques specific to our setting (such as the martingale argument used to analyze the random walk in Lemma 3.7). Another benefit of this analysis, in addition to its simplicity, is the considerably better dependence of the convergence time on the number of players and congestible resources.

Our Results and Techniques.

We show that a natural and simple multiplicative-weights algorithm indeed achieves exponential improvement over the worst correlated equilibrium, for a class of load-balancing games. Our main result is that using the Hedge algorithm [12] in the bulletin board model, the expected makespan of the outcome is bounded by $O(\log n)$, exponentially better than the known lower bounds for generic no-regret algorithms. We also show that Hedge continues to satisfy the no-regret property even in the bulletin board model.

We utilize KL-divergence to express the distance between the mixed strategy employed by a player at time t and her projected strategy at the symmetric Nash equilibrium of the non-atomic version of the game. We show that when this

[1] which is equivalent to the worst correlated equilibrium

distance is large enough, then it has a tendency to shrink. As a result we can predict the evolution of the system by analyzing a random walk, that has negative drift only when we are far away from the origin, an analysis that is of independent interest.

Prior work.

The theory of learning in games has a long history; see [13] for an extensive exposition of the literature in this field, which has primarily focused on analyzing the convergence behavior of various classes of learning processes and relating this behavior to Nash equilibrium, correlated equilibrium, and their refinements. See [6] for a more recent survey. The relationship between regret minimization, calibrated forecasting, and correlated equilibrium has been studied by [15, 16], and the connection between these topics and the price of anarchy was first made in [4, 5]. Whereas these papers use regret bounds to establish *static* equilibrium properties of the limiting distribution of play, our work requires directly analyzing the *dynamics* of the stochastic process induced by these algorithms. Adaptive procedures converging to Nash equilibrium were discovered in [18] and [14]; the latter paper is based on weak calibration, which can be interpreted as a form of regret minimization.

There has been considerable research in algorithmic game theory on understanding the behavior of adaptive procedures in load-balancing games and other congestion games, including best-response dynamics [17] and replication protocols [10]. These simple distributed protocols are well motivated, but they lack desirable learning-theoretic properties such as the no-regret property. An exception is [11], which analyzes a continuous-time process in non-atomic congestion games that can be regarded as the continuum limit of the multiplicative-weights learning process studied here. The shift from atomic to non-atomic congestion games eliminates the distinction between the solution quality of correlated, mixed Nash, and pure Nash equilibria, thus eliminating the motivating question in our work while also evading most of the technical difficulties we address in analyzing the discrete-time process in atomic congestion games. In the context of atomic congestion games, Roughgarden [24] has recently shown that for a wide class of games, including congestion games that satisfy a natural smoothness condition, bounds on the price of anarchy automatically extend to the total price of anarchy, when the global quality is defined to be the average cost.

In [19] we introduced the study of the multiplicative weights learning algorithm in atomic congestion games. Our setting was the standard full information one, where all players have access to an accurate model of the underlying game. We show that in almost all such games, the multiplicative-weights learning algorithm results in convergence to pure equilibria. As discussed earlier, shifting from the full information setting to the more realistic bulletin board model invalidates the results of [19]; in particular this shift falsifies the prediction of convergence to pure equilibria and necessitates an analysis of the dynamics using completely different tools.

2. PRELIMINARIES

A *strategic game* is a triple $(N; (S_i)_{i\in N}; (u_i)_{i\in N})$ where N is the set of players and for every player $i \in N$, S_i is the set of *(pure) strategies (or actions)* of player i, and the *utility function* u_i is a real valued function defined on $S = \times_{i\in N} S_i$. For every strategy profile $s \in S$, $u_i(s)$ represents the payoff (positive utility) to player i. For any strategy profile $s \in S$ and any strategy s'_i of player i we use (s_{-i}, s'_i) to denote the strategy profile that we derive by substituting the i-th coordinate of the strategy profile s with s'_i. A strategy profile s is a *Nash equilibrium* if $u_i(s) \geq u_i(s_{-i}, s'_i)$ for every s'_i and every $i \in N$. Analogously, a *Nash ϵ-equilibrium* is defined as a strategy profile s such that $u_i(s) \geq u_i(s_{-i}, s'_i) - \epsilon$ for every s'_i, and every $i \in N$. These notions are extended to *randomized or mixed strategies* by using the expected payoff.

External regret of an online algorithm (sometimes referred to as merely *regret*) is defined as the maximum over all input instances of the expected difference in payoff between the algorithm's actions and the best action. If this difference grows sublinearly with time, then the algorithm is said to exhibit *no external regret*, or simply *no regret*. No-regret algorithms are closely related to a notion of correlated equilibrium. We say that a probability distribution π over the strategy profiles S is a *coarse (weak) correlated equilibrium* if for all players i and strategies $s'_i \in S_i$, the expected payoff of player i playing s'_i is no better than the expected payoff from the distribution, i.e., $\sum_{s\in S} u_i(s)\pi(s) \geq \sum_{s\in S} u_i(s_{-i}, s'_i)\pi(s)$. Note that a mixed Nash equilibrium is a coarse correlated equilibrium where the probability distribution π is a product distribution over the strategy sets S_i (i.e., it is not correlated). It is well known that the long-run average outcome of repeated play using no-external regret algorithm converges to the set of coarse correlated equilibria. Similarly one can use a somewhat more restrictive notion of no-internal-regret algorithm, whose long-run average outcome of repeated play converges to the set of correlated equilibria.

3. SYSTEM ANALYSIS

In this section we study the performance of learning algorithms in load-balancing games, i.e. congestion games on parallel links using the "bulletin board model" in which players assess edge costs according to the actual cost incurred on that edge, and not the hypothetical cost if the player had used it. We demonstrate that using the Hedge algorithm in the "bulletin board model" the process remains close to the symmetric fully mixed equilibrium of the non-atomic version of the game. As a result, its performance is exponentially better than the worst correlated equilibrium of the game.

In this section we first present the definition of the games we will be focusing on (section 3.1). Next, we introduce the multiplicative updates algorithm and the bulletin board model in section 3.2, where we prove that the no-regret property persists in the bulletin board model. The main part of the analysis is in section 3.3, while we defer a few technical lemmas to section 3.4.

3.1 Defining the game and the social cost

The congestion game we consider in this section is an atomic congestion game with a set of n players, each having weight $w_i = 1/n$, and n edges with cost functions $c_e(x)$. In each period $t = 1, 2, \ldots$, each player chooses one edge e. We define $f_t(e)$ to be the total amount of flow on edge e in period t, i.e. $f_t(e) = j/n$ where j is the number of players choosing e in period t. We make the following standing assumptions: for the edge e, the function $c_e(x)$ is twice continuously differentiable, satisfies $c_e(0) = 0$ and $c_e(1) \leq 1$, and for some positive constants A, B it satisfies $c'_e(x) \geq A$ and $0 \leq c''_e(x) \leq B$ for all $x \in [0, 1]$. In section 3.4, lemma 3.8 proves that these hypotheses imply the following inequalities for all $x \in [0, 1]$:

$$Ax \leq c_e(x) \leq (B + 1)x \tag{1}$$

As a measure of social cost, we adopt the maximum edge cost, $\max_e c_e[f_e(t)]$. Interpreting players as jobs and edges as machines, this interpretation of the social cost is equivalent to the makespan. The inequality $Ax \leq c_e(x) \leq (B+1)x$ implies that for any flow vector f the social cost $\max_e c_e(f_e)$ lies between $A\|f\|_\infty$ and $(B + 1)\|f\|_\infty$. In particular, the social optimum is $\Theta(1/n)$. As we have mentioned in the introduction, even for the extremely simple case in which $c_e(x) = x$ for all e, x — i.e., a load-balancing game in which players schedule n jobs on n machines, and the cost experienced by player i is proportional to the number of jobs on its machine — the correlated equilibria of the game can be exponentially worse than any Nash equilibrium.

3.2 The learning algorithm and the bulletin board model

To define the learning algorithm used by each player, we let ε be a small positive number (we'll need to have $\varepsilon \leq 1/n^3$ for the analysis) and we introduce the following notations.

$$c_e[t] = c_e(f_t(e)), \quad c_e[1 : t] = \sum_{r=1}^{t} c_e[r]$$

$$Z(t) = \sum_{e \in E} \exp(-\varepsilon c_e[1 : t - 1]).$$

In period t, each player samples a random edge e with probability

$$P(e, t) = \frac{\exp(-\varepsilon c_e[1 : t - 1])}{Z(t)}, \tag{2}$$

i.e., to obtain $P(e, t)$ from $P(e, t - 1)$ we multiply it by $\exp(-\varepsilon c_e[t - 1])$ and then renormalize all probabilities so that they sum to 1. This algorithm for specifying a mixed strategy in period t is a version of the Hedge algorithm [12], modified so that players assess edge costs according to the actual cost $c_e[t-1]$ incurred on that edge, and not the hypothetical cost $c_e(f_t(e) + 1)$ if the player had used it for players that do not use the edge in this iteration. This model, usually referred to as the bulletin board model, has therefore been used in the design of several distributed routing protocols. Using the well-known fact that Hedge itself is a no-regret learning algorithm[2] first we prove that the bulletin board variant of Hedge is also a no-regret learning algorithm.

Proposition 3.1. *The bulletin board variant of Hedge in any load-balancing game with non-decreasing cost functions retains the no-regret property.*

Proof. Hedge is known to have the no-regret property even in settings when the cost functions of the edges can vary with time. For the proof, let us consider such a setting, where the actual cost/latency of each edge at period t as $c^t_e(x^t_e)$, where x^t_e is the load of the edge in question at period t. Naturally, all cost functions c^t_e are non-decreasing functions of x_e. Now, we will define a new cost function C^t_e as follows:

$$C^t_e(x) = \begin{cases} c^t_e(x) & \text{if } x \leq x^t_e \\ c^t_e(x^t_e) & \text{otherwise.} \end{cases}$$

Let us examine what this new cost function expresses. Under these cost functions, the latency of any edge observed at time t is actually the worst possible and any further increase on the load of any edge would have no effect on its latency. If this optimistic view of the cost of the edges were actually true, then the algorithm we have proposed would perform exactly as the Hedge algorithm. Hedge is known to have the no-regret property, hence, the expected performance of the algorithm as t goes to infinity is roughly as good as that of the best edge/strategy in hindsight under this modified costs C. However, the actual cost of any strategy under the real cost functions c, when taking into account the effect of the deviating player, would be at least as bad as that under the optimistic costs C. As a result the performance of our algorithm is also of *epsilon*-regret in regards to the best strategy in hindsight under the true cost evaluations. □

Although the proposition above in its current form will suffice for our purposes, it can be straightforwardly extended to any no-regret algorithm and all congestion games with non-decreasing cost functions.

3.3 Main theorems

The main result of this section is the following bound on the distribution $P(t)$ determined by the Hedge algorithm (2).

Theorem 3.2. *If all players sample their strategies at time t using the distribution $P(t)$ determined by the Hedge algorithm (2), then there exist positive constants α, β_0 such that for all times t and all $\beta > \beta_0$ it holds with probability at least $1 - \exp(-\alpha\beta)$ that $\max_e |P(e, t)| < 2\beta/n$.*

Combining this theorem with Chernoff bounds leads to a price-of-anarchy type result — the long-run average social cost exceeds the social optimum by a factor of at most $O(\log n)$. More precisely:

Corollary 3.3. *In the setting of Theorem 3.2, there exist constants c_1, c_2 such that for all t, with probability at least $1 - 1/n^{c_1}$, the flow f_t sampled by the players satisfies*

$$\max_e c_e(f_t(e)) \leq \frac{c_2 \log n}{n}.$$

The proof of Theorem 3.2 rests on analyzing a stochastic process $KL(t)$ defined as the KL-divergence between the Nash equilibrium and $P(t)$. Let Q be the symmetric Nash equilibrium of the non-atomic congestion game (where all players play the same strategy) with edge set E and cost functions $(c_e)_{e \in E}$. KL-divergence between P and Q is defined as

$$KL(t) = \sum_{j \in E} Q(j) \log (Q(j)/P(j, t)).$$

KL-divergence measures the distance between the distributions $Q(j)$ and $P(j, t)$[3]. It is zero if they are equal and

[2]provided that ε converges to zero at an appropriate rate depending on t, e.g. $\varepsilon(t) = O(1/n^3\sqrt{t})$

[3]although it is not a true distance metric since it is not symmetric

positive otherwise. We will show that when this distance is large enough, then it has a tendency to shrink (Lemma 3.6). This reduces the analysis of $KL(t)$ to exploring the behavior of new kind of random walks, which face negative drift only when they are far away from the origin. Lemma 3.7 provides this analysis.

Theorem 3.2 will follow from proving an exponential tail bound for $KL(t)$.

THEOREM 3.4. *There exist positive constants α, β_0 such that $\Pr(KL(t) > \beta/n) < e^{-\alpha\beta}$ for all $\beta > \beta_0$.*

We next sketch the proof of this tail bound. In all of the following arguments, "log" denotes the natural logarithm function. We'll need the following technical lemma.

LEMMA 3.5.

$$\log Z(t+1) - \log Z(t) \le (\exp(-\varepsilon) - 1) \sum_e P(e,t)c(e,t).$$

PROOF. We will use the fact that if $0 \le y \le 1$, then $\exp(-\varepsilon y) \le 1 + y(\exp(-\varepsilon) - 1)$; this can be verified by checking that the left side is a convex function, the right side is a linear function, and the left and right sides are equal when y is an endpoint of the interval $[0, 1]$.

$$\begin{aligned}\frac{Z(t+1)}{Z(t)} &= \frac{\sum_e \exp(-\varepsilon c_e[1:t-1])\exp(-\varepsilon c_e[t])}{Z(t)} \\ &\le \frac{\sum_e \exp(-\varepsilon c_e[1:t-1])\,[1 + c_e[t](\exp(-\varepsilon) - 1)]}{Z(t)} \\ &= 1 + (\exp(-\varepsilon) - 1)\sum_e P(e,t)c_e[t].\end{aligned}$$

The lemma follows by taking the logarithm of both sides and using the identity $\log(1 + y) \le y$. □

We denote the difference $KL(t + 1) - KL(t)$ as Δ_t.

LEMMA 3.6. *The stochastic process $KL(t)$ satisfies*

$$\mathbf{E}[\Delta_t \mid P(t)] \le -(AC\varepsilon/n)KL(t) + C\varepsilon/n^2. \quad (3)$$

In particular, $KL(t)$ drifts to the left at a rate of $\Omega(\varepsilon/n^2)$ whenever it is greater than $2/(An)$.

PROOF. A simple calculation using equation (2) using Lemma 3.5 justifies the bound

$$\log P(e,t) - \log P(e,t{+}1) \le \varepsilon c_e[t] - \left(1 - e^{-\varepsilon}\right)\sum_{e'} P(e',t)c_{e'}[t]. \quad (4)$$

Taking a weighted average of the bounds (4), weighted by $Q(e)$,

$$\begin{aligned}\Delta_t &= \sum_e Q(e)\left(\log P(e,t) - \log P(e,t+1)\right) \\ &\le \varepsilon \sum_e Q(e)c_e[t] - (1 - \exp(-\varepsilon))\sum_e P(e,t)c_e[t].\end{aligned}$$

Now, using $\bar{c}_e[t]$ to denote $\mathbf{E}[c_e[t] \mid P(t)]$ and using $c_e[\bar{f}(t)]$ denote $c_e(P(e,t)/n)$, we may take the conditional expectation of both sides and apply the identity $1 - \exp(-\varepsilon) > \varepsilon - \frac{1}{2}\varepsilon^2$ to obtain:

$$\begin{aligned}\mathbf{E}\left[\Delta_t \mid P(t)\right] &\le \varepsilon \sum_e [Q(e) - P(e,t)]\bar{c}_e[t] + \frac{\varepsilon^2}{2}\sum_e P(e,t)\bar{c}_e[t] \\ &\le \varepsilon(Q - P(t)) \cdot c[\bar{f}(t)] + \varepsilon(Q - P(t)) \cdot (\bar{c}[t] - c[\bar{f}(t)]) + \frac{\varepsilon^2}{2}.\end{aligned}$$

We denote the usual convex potential function $\sum_e \int_0^{x_e} c_e(y)\,dy$ as $\Phi(\vec{x})$. As a result, we have for the first term above that

$$\varepsilon(Q - P(t)) \cdot \nabla\Phi(P(t)) \le \varepsilon\left[\Phi(Q) - \Phi(P(t))\right] \le -A\varepsilon\|P(t) - Q\|_2^2,$$

where the last inequality uses the fact the Q minimizes Φ, combined with our assumption that $c'_e(y) \ge A$ for all y. It is not hard to prove that for some constant C, the additional inequalities

$$\begin{aligned}\|P(t) - Q\|_2^2 &\ge \frac{C}{n}KL(t), \quad (5) \\ \varepsilon(Q - P(t)) \cdot (\bar{c}[t] - c[\bar{f}(t)]) + \frac{1}{2}\varepsilon^2 &\le C\varepsilon/n^2 \quad (6)\end{aligned}$$

hold (Lemmas 3.11 and 3.12 in section 3.4), implying that the stochastic process $KL(t)$ satisfies

$$\mathbf{E}[\Delta_t \mid P(t)] \le -(AC\varepsilon/n)KL(t) + C\varepsilon/n^2, \quad (7)$$

as claimed. □

Next we give the submartingale argument to show that the fact that $KL(t)$ has negative drift when its large implies that the probability of $KL(t) > \beta/n$ is exponentially small in β as claimed by Theorem 3.4.

LEMMA 3.7. *Let $(Y_t)_{t\ge 0}$ be a random walk satisfying the following for some constant $M \ge 1$:*

bounded differences: $|Y_{t+1} - Y_t| \le 1$;
negative drift: $\mathbf{E}(Y_{t+1} - Y_t \mid Y_t) \le -1/M$ *whenever $Y_t \ge M$.*

Then there exist constants α, λ_0 such that for all $\lambda > \lambda_0$, we have $\Pr(Y_t > \lambda M) < e^{-\alpha\lambda}$.

PROOF. Let $Z_s = Y_s + s/M$. For any $t \ge 0$ and $\gamma \ge 1$, we may condition on the event $Y_t \ge 2\gamma M$ thus defining a new probability space $\Omega(t, \gamma)$ consisting only of the sample points where this event holds, with their probabilities scaled so that they sum to 1. The random variables $\{Z_s \mid t \le s \le t + \gamma M\}$ constitute a supermartingale on $\Omega(t, \gamma)$, because

$$\mathbf{E}(Z_{s+1}|Z_s) = \frac{s+1}{M} + \mathbf{E}(Y_{s+1}|Y_s) \le \frac{s+1}{M} - \frac{1}{M} + Y_s = Z_s,$$

the inequality being justified because $Y_s \ge M$ for all $s \in \{t, t+1, \ldots, t+\gamma M.\}$ The supermartingale satisfies $|Z_{s+1} - Z_s| \le 1 + \frac{1}{M} \le 2$, hence we may apply Azuma's inequality to conclude that

$$\begin{aligned}\Pr(Z_{t+\gamma M} - Z_t > \gamma M/2) &< \exp\left(-\frac{\gamma^2 M^2}{4} \cdot \frac{1}{8\gamma M}\right) \\ &= e^{-\gamma M/32}.\end{aligned}$$

Rewriting the left side in terms of $Y_{t+\gamma M}$ and Y_t, we see that the inequality expresses an upper bound on $\Pr(Y_{t+\gamma M} > Y_t - \gamma M/2)$. Now for any $t, \lambda \ge 0$ let $R(t, \lambda) = \Pr(Y_t - Y_0 > \lambda M)$. We will prove, by induction on t, that $\lambda > 320\ln(2)$ implies $R(t, \lambda) < e^{-\lambda M/80}$, thus establishing the lemma. For $t = 0$ the claim holds trivially. Otherwise, let $L = \lambda M/4$, let $u = \max\{t - 2L, 0\}$, and let $\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3$ denote the events $\{Y_u < 4L\}, \{4L \le Y_u < 5L\}, \{5L \le Y_u\}$ respectively. We have

$$\begin{aligned}&\Pr(Y_t > \lambda M) = \Pr(Y_t > 4L \mid \mathcal{E}_1)\Pr(\mathcal{E}_1) + \\ &+ \Pr(Y_t > 4L \mid \mathcal{E}_2)\Pr(\mathcal{E}_2) + \Pr(Y_t > 4L \mid \mathcal{E}_3)\Pr(\mathcal{E}_3) \\ &\le 0 + e^{-L/16} + R(u, 5\lambda/4) \le 2e^{-\lambda M/64} < e^{-\lambda M/80},\end{aligned}$$

the final inequality following from our assumptions that $\lambda > 320\ln(2)$ and $M \ge 1$. □

Proof of Theorems 3.2 and 3.4:

Let $Y_t = KL(t)/\varepsilon$ and apply Lemma 3.7 with $M = (A + C)n^2/AC$. The inequality $\mathbf{E}[\Delta_t \mid P(t)] \le -(AC\varepsilon/n)KL(t) + C\varepsilon/n^2$ implies that there exist positive constants α, β_0 such that $\Pr(KL(t) > \beta/n) < e^{-\alpha\beta}$ for all $\beta > \beta_0$. This proves Theorem 3.4. The bound on $\max_e |P(e,t)|$ in Theorem 3.2 now follows by combining the KL-divergence bound in Theorem 3.4 with Lemma 3.9 below, which bounds the infinity-norms of two distributions P, Q in terms of their corresponding KL-divergence. $\square$

3.4 Technical lemmas

The following technical lemmas complete the analysis the performance of Hedge:

Lemma 3.8. *Let $c_e(x)$ be a function in $C^2([0,1])$ satisfying*

- $c_e(0) = 0, c_e(1) \le 1$;
- *for all* $x \in [0,1]$, $c_e'(x) \ge A$;
- *for all* $x \in [0,1]$, $0 \le c_e''(x) \le B$.

Then $Ax \le c_e(x) \le A(B+1)x$ for all $x \in [0,1]$.

Proof. For all x we have $c_e(x) = \int_0^x c_e'(y)\,dy \ge \int_0^x A\,dy$, which establishes that $Ax \le c_e(x)$. To establish the upper bound on $c_e(x)$, we first use the mean value theorem to deduce that there exists some $x \in [0,1]$ such that

$$c_e'(x) = \frac{c_e(1) - c_e(0)}{1-0} \le 1.$$

If there exists $y \in [0,1]$ such that $c_e'(y) > B+1$, then a second application of the mean value theorem would imply the existence of $z \in [0,1]$ satisfying

$$|c_e''(z)| = \left| \frac{c_e'(y) - c_e'(x)}{y-x} \right| > B,$$

contradicting our hypothesis about c_e. Hence $c_e'(y) \le B+1$ for all $y \in [0,1]$. Now, for all $x \in [0,1]$, $c_e(x) = \int_0^x c_e'(y)\,dy \le \int_0^x B+1\,dy$, which establishes that $c_e(x) \le (B+1)x$. $\square$

Lemma 3.9. *If P, Q are two probability distributions on a finite set S, satisfying $\|P\|_\infty \ge 2\|Q\|_\infty$, then $KL(Q;P) \ge \frac{\|P\|_\infty}{16}$.*

Proof. Let s_0 be a point at which $P(s_0) = \|P\|_\infty$. Let $a = Q(s_0), b = P(s_0)$. Then

$$\begin{aligned} KL(Q;P) &= Q(s_0)\log(\frac{Q(s_0)}{P(s_0)}) + \sum_{s \ne s_0} Q(s)\log(\frac{Q(s)}{P(s)}) \\ &= a\log(a/b) + (1-a)\sum_{s\ne s_0} \frac{Q(s)}{1-a}\left[-\log\frac{P(s)}{Q(s)}\right]. \end{aligned}$$

Since $\sum_{s \ne s_0} Q(s)/(1-a) = 1$, the sum on the right side can be interpreted as a weighted average of value of the convex function $-\log(x)$ at the points $P(s)/Q(s)$. Using Jensen's inequality, we see that this is greater than or equal to $-\log(x)$ evaluated at the point

$$x = \sum_{s \ne s_0} \left(\frac{Q(s)}{1-a}\right)\frac{P(s)}{Q(s)} = \sum_{s\ne s_0}\frac{P(s)}{1-a} = \frac{1-b}{1-a}.$$

Hence we have derived the first line of the following series of bounds.

$$KL(Q;P) \ge a\log\left(\frac{a}{b}\right) + (1-a)\left(\frac{1-a}{1-b}\right) = \int_a^b \frac{x-a}{x(1-x)}\,dx.$$

The integrand is a strictly increasing function of x for $0 < x < 1$, so letting $c = (a+b)/2$ we have

$$\begin{aligned} \int_a^b \frac{x-a}{x(1-x)}\,dx &\ge \int_c^b \frac{x-a}{x(1-x)}\,dx \\ &\ge (b-c)\frac{c-a}{c(1-c)} = \frac{1}{4}\frac{(b-a)^2}{c(1-c)} \\ &\ge \frac{(b-a)^2}{4b}. \end{aligned}$$

The assumption that $\|P\|_\infty \ge 2\|Q\|_\infty$ implies $a \le b/2$, and the lemma follows immediately. $\square$

Lemma 3.10. *In a non-atomic parallel-links congestion game with n edges whose cost functions satisfy the conditions of Lemma 3.8, the Nash equilibrium Q satisfies for every edge e:*

$$\frac{A}{(B+1)n} \le Q(e) \le \frac{B+1}{An}$$

Proof. Since Q is a Nash equilibrium, there exists a $z > 0$ such that $c_e(Q(e)) = z$ for all e. Since there is some e_0 such that $Q(e_0) \le 1/n$, we have $z \le (B+1)/n$. Now for any edge e, the relations $AQ(e) \le c_e(Q(e))$ and $c_e(Q(e)) = z \le (B+1)/n$ together imply that $Q(e) \le (B+1)/(An)$. Similarly, the existence of an edge e_1 such that $Q(e_1) \ge 1/n$ implies that $z \ge A/n$ from which it follows that $Q(e) \ge A/((B+1)n)$ for all e. $\square$

Lemma 3.11. *For any distributions P, Q on an n-element set S, if $C/n \le Q(s) \le 1/2$ for all $s \in S$, then*

$$\|P - Q\|_2^2 \ge \frac{C}{n} KL(Q;P).$$

Proof. Let $x(s) = P(s) - Q(s)$. We have

$$\begin{aligned} KL(Q;P) &= -\sum_s Q(s)\log\frac{P(s)}{Q(s)} \\ &= -\sum_s Q(s)\log\left(1 + \frac{x(s)}{Q(s)}\right). \end{aligned}$$

Using the identity $\log(1+x) \ge x - x^2$, valid for $-1/2 \le x \le 1$, we obtain

$$\begin{aligned} KL(Q;P) &\le -\sum_s Q(s)\left[\frac{x(s)}{Q(s)} - \frac{x(s)^2}{Q(s)^2}\right] \\ &\le \sum_s \frac{x(s)^2}{Q(s)} \le \frac{n}{C}\|x\|_2^2, \end{aligned}$$

from which the lemma follows immediately. $\square$

Lemma 3.12. *Let P be any probability distribution on edges and let $f = (f_e)_{e\in E}$ be the random flow vector obtained by letting n players each sample an edge in E according to P and send $1/n$ units of flow on that edge. Let $\bar{\mathbf{c}}, \mathbf{c}$ denote the vectors*

$$\bar{\mathbf{c}}_e = \mathbf{E}(c_e(f_e)), \qquad \mathbf{c}_e = c_e(\mathbf{E}(f_e)) = c_e(P(e)),$$

respectively. There is a constant C such that

$$\varepsilon(Q-P)\cdot(\bar{\mathbf{c}}-\mathbf{c})+\frac{1}{2}\varepsilon^2 \le \frac{C\varepsilon}{n^2}.$$

PROOF. Let us fix our attention on one edge e and let $x_0 = P(e)$. Taylor's theorem with remainder ensures that for all $x \in [0,1]$,

$$c_e'(x_0)(x-x_0) \le c_e(x)-c_e(x_0) \le +c_e'(x_0)(x-x_0)+\frac{B}{2}(x-x_0)^2,$$

since $0 \le c_e''(y) \le B$ for all y. Plugging the random variable f_e into this bound, we find that

$$\mathbf{c}_e \le \bar{\mathbf{c}}_e \le \mathbf{c}_e + c_e'(x_0)\mathbf{E}(f_e - x_0) + \frac{B}{2}\mathbf{E}((f_e-x_0)^2)$$

$$0 \le \bar{\mathbf{c}}_e - \mathbf{c}_e \le \frac{B}{2}\,\mathrm{Var}(f_e).$$

If z_i $(i = 1,2,\ldots,n)$ denotes a collection of independent Bernoulli random variables with $\Pr(z_i = 1) = P(e)$, then the random variable f_e has the same distribution as $\frac{1}{n}\sum_{i=1}^{n} z_i$, so its variance is

$$\mathrm{Var}(f_e) = \frac{1}{n^2}\cdot n \cdot \mathrm{Var}(z_i) = \frac{P(e)(1-P(e))}{n}.$$

To bound the dot product $(Q-P)\cdot(\bar{\mathbf{c}}-\mathbf{c})$ from above, we first note that when $Q(e) < P(e)$ we have $(Q(e)-P(e))(\bar{\mathbf{c}}_e-\mathbf{c}_e) \le 0$. The remaining terms of the dot product according to lemma 3.10 satisfy $Q(e)-P(e) \le Q(e) \le (B+1)/(An)$, and $P(e)(1-P(e))/n \le P(e)/n \le Q(e)/n \le (B+1)/(An^2)$. Hence the dot product is bounded above by $\sum_e \frac{B+1}{An}\cdot\frac{B+1}{An^2} = \left(\frac{B+1}{An}\right)^2$. Recalling that $\varepsilon \le 1/n^2$, we see that the inequality in the statement of the lemma is satisfied by setting $C = \frac{1}{2}+\left(\frac{B+1}{A}\right)^2$. $\square$

4. SUMMARY

Given that online learning is quite thoroughly understood in the setting of a single learner [3], it is rather natural to hope for a thorough understanding of systems consisting of multiple learners, but the characterization of such systems in existing work is far from thorough. Several recent papers have pursued this direction in the context of no-regret learning [4, 5, 24], but their findings have been limited to games in which the no-regret property by itself suffices to establish bounds on the overall system performance. Our work establishes that in many cases of interest — and specifically in settings closely related to the reality of distributed systems — this optimistic view does not materialize. Two systems consisting of no-regret learners can exhibit huge performance differences. Nevertheless, our result is in essence a positive result. It shows that "natural" candidates (e.g. Hedge) of no-regret algorithms perform well. An interesting direction for future research is the question of how much we can extend the family of allowable no-regret algorithms while still allowing for strong provable performance bounds on the overall system behavior.

5. REFERENCES

[1] R. Aumann. *Subjectivity and correlation in randomized strategies*. Journal of Mathematical Economics 1:67-96, 1974.

[2] Baruch Awerbuch, Rohit Khandekar, Satish Rao. *Distributed algorithms for multicommodity flow problems via approximate steepest descent framework*. SODA 2007: 949-957

[3] N. Cesa-Bianchi and G. Lugosi. *Prediction, learning, and games*. Cambridge University Press, 2006

[4] A. Blum, E. Even-Dar, and K. Ligett, *Routing without regret: on convergence to Nash equilibria of regret-minimizing algorithms in routing games*. Proceedings of the twenty-fifth annual ACM symposium on Principles of distributed computing, ACM Press, 45-52, 2006.

[5] A. Blum, M. Hajiaghayi, K. Ligett, and A. Roth, *Regret Minimization and the Price of Total anarchy*. STOC 2008, pages 373-382.

[6] A. Blum and Y. Mansour, *Learning, Regret Minimization and Equilibria*, in *Algorithmic Game Theory* N. Nisan, T. Roughgarden, E. Tardos and V. Vazirani (eds), Cambridge University Press, 2007.

[7] X. Chen and X. Deng. Settling the Complexity of *Two-Player Nash Equilibrium*. FOCS, 2006.

[8] A. Czumaj and B. Vöcking. *Tight Bounds for Worst-Case Equilibria*. SODA, 2002; pp. 413-420.

[9] C. Daskalakis, P. Goldberg, and C. Papadimitriou. *The Complexity of Computing a Nash Equilibrium*. STOC, 2006.

[10] S. Fischer, H. Räcke, and B. Vöcking. *Fast convergence to Wardrop equilibria by adaptive sampling methods*. STOC, 2006.

[11] S. Fischer and B. Vöcking. *On the evolution of selfish routing*. ESA, 2004.

[12] Y. Freund, and R. Schapire. *Adaptive game playing using multiplicative weights*. Games and Economic Behavior, 29:79-103, 1999.

[13] D. Fudenberg, and D. Levine. *The Theory of Learning in Games*. MIT Press. 1998.

[14] D. Foster and S. M. Kakade *Deterministic Calibration and Nash Equilibrium*. In Proceedings of Conference on Learning Theory (COLT), 2004.

[15] D. Foster and R. Vohra. *Regret in the on-line decision problem*. Games and Economic Behavior, 21:40-55, 1997

[16] D. Foster, and R. Vohra. *Calibrated learning and correlated equilibrium*. Games and Economic Behavior, 1997.

[17] M. Goemans, V. Mirrokni, and A. Vetta, *Sink Equilibria and Convergence*. FOCS, 2005.

[18] S. Hart, and A. Mas-Colell. *Stochastic Uncoupled Dynamics and Nash Equilibrium*. Games and Economic Behavior 57, 286-303, 2006.

[19] R. Kleinberg and G. Piliouras and E. Tardos. *Multiplicative updates outperform generic no-regret learning in congestion games*. STOC, 2009.

[20] E. Koutsoupias and C. H. Papadimitriou. *Worst-case equilibria*. In Proceedings of the 16th Annual Symposium on Theoretical Aspects of Computer Science, 404-413, 1999.

[21] M. Mavronicolas and P. Spirakis. *The price of selfish routing*. STOC, 2001; pages 510-519.

[22] V. Mirrokni and A. Vetta. *Convergence Issues in Competitive Games*. RANDOM-APPROX, 2004.

[23] Michael Mitzenmacher. *How useful is old information?*. In Proc. 16th Ann. ACM SIGACT-SIGOPS Symp. on Principles of Distributed Computing (PODC), pages 83-91, 1997.

[24] T. Roughgarden. *Intrinsic Robustness of the Price of Anarchy*. STOC, 2009.

[25] A. Tang, J. Wang, S. Low. *Counter-Intuitive Behaviors in Networks under End-to-end Control*. IEEE /ACM Transactions on Networking, 14(2):355-368, April 2006.

[26] H. P. Young, *Strategic Learning and Its Limits*. Arne Ryde Memorial Lectures. Oxford, UK: Oxford University Press, 2004.