Finite-time convergence to an ϵ-efficient Nash equilibrium in potential games

Anna Maddux [*1] **Reda Ouhamma** [*1] **Maryam Kamgarpour** [1]

Abstract

This paper investigates the convergence time of log-linear learning to an ϵ-efficient Nash equilibrium (NE) in potential games. In such games, an efficient NE is defined as the maximizer of the potential function. Existing results are limited to potential games with stringent structural assumptions and entail exponential convergence times in $1/\epsilon$. Unaddressed so far, we tackle general potential games and prove the first finite-time convergence to an ϵ-efficient NE. In particular, by using a problem-dependent analysis, our bound depends polynomially on $1/\epsilon$. Furthermore, we provide two extensions of our convergence result: first, we show that a variant of log-linear learning that requires a factor A less feedback on the utility per round enjoys a similar convergence time; second, we demonstrate the robustness of our convergence guarantee if log-linear learning is subject to small perturbations such as alterations in the learning rule or noise-corrupted utilities.

1. Introduction

Interactions of multiple agents are at the heart of many applications in transportation networks, auctions, telecommunication networks, and multi-robot systems. In game theory, the Nash equilibrium is a popular solution concept to describe outcomes of a multi-agent system (Nash, 1951). Fundamental considerations in game theory are whether an NE exists, if strategic players can learn it, and if so at which speed they can learn it. Furthermore, for practical purposes it is important to understand which NE is learned. This is particularly pertinent in games that admit a social welfare function, as it enables the definition of an efficient NE as an NE that maximizes the social welfare. The social welfare is typically an aggregate measure of individual utilities such as their sum or a measure of fairness. In distributed control of engineering systems, for example, studying efficient NEs is important, as the aim is to optimize some global objective function in a distributed manner.

The class of potential games (Monderer and Shapley, 1996b) lends itself to studying efficient NEs since every joint action maximizing the potential corresponds to an NE. This property follows from the fact that in potential games, the difference in a player's utility generated by a unilateral change of her action equals the difference in potential. Consequently, if the social welfare function is aligned with the potential function, meaning that an increase in social welfare is associated with an increase in potential (Paccagnan et al., 2022), then any joint action maximizing the potential is an efficient NE (Marden and Shamma, 2015). In identical interest games, for example, maximizing the aggregated utility is achieved by maximizing the potential function which is trivially given by the common utility function. In the example of coverage problems, the goal is to assign players to locations to achieve maximal coverage. When each player's utility is designed as the marginal contribution then this goal is achieved by maximizing the resulting potential function (Marden and Wierman, 2013).

In this paper, we investigate the speed at which an efficient NE can be reached in potential games.

1.1. Related work

In potential games, several learning rules exist such as iterative best-response dynamics (Rosenthal, 1973; Awerbuch et al., 2008; Chien and Sinclair, 2011), no-regret algorithms (Krichene et al., 2015; Palaiopanos et al., 2017; Heliou et al., 2017), and fictitious play (Monderer and Shapley, 1996a;b) for which asymptotic convergence to an NE is guaranteed. However, only log-linear learning (Blume, 1993; Young, 1993) and variants thereof (Arslan et al., 2007; Marden et al., 2007; Marden and Shamma, 2012) are known to converge to an efficient Nash equilibrium. In log-linear learning, the players asynchronously choose an action with a probability proportional to its exponentiated utility. To increase its applicability, existing works also studied log-linear learning under more realistic structural assumptions. For example, Leslie and Marden (2011) proves that log-linear learning also handles the practical setting where the observed utilities

[*]Equal contribution [1]SYCAMORE Lab, École Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland. Correspondence to: Anna Maddux <anna.maddux@epfl.ch>, Reda Ouhamma <reda.ouhamma@epfl.ch>.

Workshop on Foundations of Reinforcement Learning and Control at the 41^{st} International Conference on Machine Learning, Vienna, Austria.

are corrupted by noise. Furthermore, Arslan et al. (2007) proposes binary log-linear learning, a slight modification of log-linear learning that only requires two points of feedback per round. Although the above works established asymptotic convergence of log-linear learning to an efficient NE, a finite-time analysis is missing for general potential games.

Few past works provide finite-time guarantees for log-linear learning to an ϵ-efficient NE, an action profile whose potential is ϵ-close to its maximum value. For instance, for atomic routing games with polynomial costs of degree at most p, Asadpour and Saberi (2009) proves a convergence time exponential in p and $1/\epsilon$ and polynomial in N, where N denotes the number of players. Moreover, Montanari and Saberi (2008; 2010) study games with graph structures between players and prove an exponential convergence time in N and $1/\epsilon$ in the worst case. Finally, in potential games with interchangeable players and a Lipschitz-continuous potential function, Shah and Shin (2010) shows a convergence time exponential in A and $1/\epsilon$ and linear in N, where A is the number of actions per player.

For the class of games we address in this paper, namely general potential games, an exponential dependence on N in the convergence time is unavoidable. In fact, finding an ϵ-efficient NE is equivalent to finding an ϵ-optimizer of the potential function, which is NP-complete (Burer and Letchford, 2012). However, there is no hardness result that justifies the exponential dependence on $1/\epsilon$ present in previous works. In this paper, we focus on deriving the first finite-time convergence guarantees of log-linear learning that hold for general potential games; are polynomial in $1/\epsilon$; and remain valid under relaxed structural assumptions such as limited access to feedback and noisy utilities.

1.2. Contributions

We study the convergence time of log-linear learning to an ϵ-efficient Nash equilibrium in general potential games. Our contributions are summarized as follows:

- We prove a convergence time of $\tilde{\mathcal{O}}((A^N/\epsilon)^{\frac{1}{\max\{\epsilon,\Delta\}}})$ to an ϵ-efficient NE (Theorem 3.1), where Δ is a problem-dependent constant. If in addition, the players are interchangeable, then an ϵ-NE is reached in $\tilde{\mathcal{O}}((\frac{N^A}{\epsilon})^{\frac{1}{\max\{\epsilon,\Delta\}}})$ which in contrast to general potential games is polynomial in N as well (Corollary B.3).
- We consider two variants of log linear learning: binary log-linear learning (Theorem 4.1) and perturbed log-linear learning (Theorem 4.3) motivated by limited feedback and noise corrupted utilities, respectively. For these variants we show convergence guarantees that are polynomial in $1/\epsilon$.

On the technical side, in their convergence analysis, most past works leveraged that log-linear learning induces a Markov chain. To obtain our novel finite-time results we build on this connection and develop new Markov chain results that can be summarized as follows:

- We derive improved mixing time bounds for a given class of Markov chains based on log-Sobolev inequalities (Lemma 2.2). In particular, this broad class of Markov chains includes those induced by log-linear learning and binary-log linear learning.
- We derive a tight Lipschitz constant for the known result regarding the Lipschitz-continuity of stationary distributions of Markov chains as a function of their transition matrix (Lemma 4.2). We leverage this result to study the convergence of learning rules such as perturbed log-linear learning for which the explicit stationary distributions are unknown (Theorem 4.3).

Notations We denote by $[N]$ the set $\{1,\dots,N\}$. For a finite set $\mathcal{X}$, we denote by $\Delta(\mathcal{X})$ the probability simplex over $\mathcal{X}$, and by $\mathbb{1}_{a\in\mathcal{X}}$ the indicator function of $\mathcal{X}$. Finally, we use the big-$\mathcal{O}$ notations $\tilde{\mathcal{O}}$ and $\tilde{\Omega}$ to hide logarithmic terms.

2. Preliminaries

2.1. Problem setup

We consider a repeated potential game with N players. Every player has an action set $\mathcal{A}$ of cardinality $A<\infty$, which for simplicity we assume to be the same for all players. The utility of player i is a mapping $U_i:\mathcal{A}^N\to[0,1]$ from the joint action space $\mathcal{A}^N$ to $[0,1]$. In a potential game, the utility functions are characterized by a potential function $\Phi:\mathcal{A}^N\to\mathbb{R}$ such that for all $i\in[N]$, $a_i,a_i'\in\mathcal{A}$, and $a_{-i}\in\mathcal{A}^{N-1}$, it holds that:

$$U_i(a_i,a_{-i})-U_i(a_i',a_{-i})=\Phi(a_i,a_{-i})-\Phi(a_i',a_{-i}).$$

In this paper, we assume that an initial action profile is drawn from a distribution μ^0. The potential game is then repeated over multiple rounds. We assume a turn-based setting, where at round t, a player $i\in[N]$ is uniformly selected to update her action while the other players remain with their previous action, *i.e.,* $a_{-i}^t=a_{-i}^{t-1}$.

A common assumption in game theory is that the players are rational, *i.e.,* that they seek to maximize their utility. Under this assumption, a natural solution concept is the pure Nash equilibrium (Nash Jr, 1950).

Definition 2.1 (Nash equilibrium)**.** A pure Nash equilibrium is an action profile $(a_i^{NE})_{i\in[N]}\in\mathcal{A}^N$ such that every player is playing a best response to the other players' actions a_{-i}^{NE}, *i.e.,*

$$U_i(a_i,a_{-i}^{NE})\le U_i(a_i^{NE},a_{-i}^{NE}),\quad \forall i\in[N],\forall a_i\in\mathcal{A},$$

where we define $a_{-i} := (a_j)_{j\in[N]\setminus\{i\}}$ as the joint action of all players except i.

In a pure NE, no player can improve her utility by unilaterally changing her action. Hereafter, we simply refer to a pure Nash equilibrium as a Nash equilibrium. A stronger solution concept specific to potential games is the efficient Nash equilibrium. An efficient Nash equilibrium is an action profile a^* which maximizes the potential function, *i.e.,* $a^* \in \arg\max_{a\in A^N} \Phi(a)$. An appealing property of finite potential games is that a potential maximizer exists, and thus an efficient NE exists (Monderer and Shapley, 1996b).

Practically, each player selects her actions based on some learning rule. In this work, we are interested in learning rules that converge, in expectation, to an ϵ-efficient NE, *i.e.,* an action profile maximizing the potential up to an additive constant:

$$\mathbb{E}[\Phi(a)] \geq \max_{a\in\mathcal{A}^N} \Phi(a) - \epsilon,$$

where the randomness stems from the distribution of action profiles at round t. The number of rounds needed to find an ϵ-efficient NE denotes the convergence time.

Connection to Markov chains: If all players apply a learning rule that relies exclusively on the utility given the most recent action profile of the other players then the considered game induces a Markov chain over the state space $\mathcal{A}^N$. In particular, the state of the Markov chain at time t corresponds to the action profile at round t and the learning rule of each player specifies the transition dynamics. This connection between the learning dynamics in potential games and Markov chains has been exploited in previous works (Blume, 1993; Young, 1993; Marden and Shamma, 2012; Shah and Shin, 2010) and is crucial for our subsequent convergence analysis. We provide the necessary background on Markov chains in the next section.

2.2. Background on Markov chains

We briefly review concepts and properties of Markov chains used throughout this paper. Consider a time-homogeneous Markov chain $\{X_t\}_{t\in\mathbb{N}}$ over the state space $\mathcal{A}^N$ with a transition matrix $P \in \mathbb{R}^{A^N\times A^N}$. The ergodic theorem (Levin and Peres, 2017) states that if a Markov chain $\{X_t\}_{t\in\mathbb{N}}$ is irreducible and aperiodic, then it has a unique stationary distribution μ, and from any initial distribution μ^0 the distribution $\mu^t = \mu^0 P^t$ converges to μ. The convergence time to the stationary distribution is quantified by the mixing time:

$$t^P_{\text{mix}}(\epsilon) := \min\{t \in \mathbb{N} \mid \|\mu^t - \mu\|_{TV} \leq \epsilon\}, \qquad (1)$$

where the total variation distance is defined as $\|\mu^t - \mu\|_{TV} := \frac{1}{2}\sum_{a\in\mathcal{A}^N} |\mu^t(a) - \mu(a)|$. Based on a remark in (Diaconis and Saloff-Coste, 1996, Section 3) we now derive a bound on the mixing time of Markov chain $\{X_t\}_{t\in\mathbb{N}}$.

Lemma 2.2. *If P is irreducible and aperiodic, then the mixing time has the following upper bound:*

$$t^P_{\text{mix}}(\epsilon) \leq \frac{1}{\rho(PP^*)}\left(\log\log\frac{1}{\mu_{\min}} + 2\log\frac{1}{\epsilon}\right), \qquad (2)$$

where $\rho(PP^*)$ denotes the log-Sobolev constant of PP^* defined in Equation (10) in Appendix A, $\mu_{\min} := \min_{a\in A^N}\mu(a)$, and P^* is the time-reversal of P.[1]

We briefly discuss the mixing time bound above and provide a proof and a thorough discussion in Appendix A. While classical approaches commonly bound the mixing time by the spectral gap defined in Equation (11) in Appendix A, bounds using log-Sobolev constants are often significantly tighter. Indeed, the mixing time upper bound using the log-Sobolev constant grows as $\mathcal{O}(\log\log(1/\mu_{\min}))$ whereas bounds using the spectral gap grow as $\mathcal{O}(\log(1/\sqrt{\mu_{\min}}))$. However, unlike the spectral gap deriving log-Sobolev constants can be extremely difficult. So constants have not been well-explored. In the next section, we derive a novel bound on the log-Sobolev constant of a class of Markov chains.

3. Log-linear learning

In this section, we review the well-established log-linear learning rule (Blume, 1993) and state our main theoretical result on the convergence time of log-linear learning to an ϵ-efficient Nash equilibrium. Our convergence analysis relies on the mixing time bound from Lemma 2.2 of the previous section.

3.1. Algorithm and background

We assume that all players follow the log-linear learning rule which is repeated over several rounds. At round t a player denoted by i is randomly chosen among all players and allowed to alter her action while the other players repeat their current action, *i.e.,* $a^t_{-i} = a^{t-1}_{-i}$. Given full-information feedback, player i observes her utility for all actions $a_i \in \mathcal{A}$ given the other players' actions a^{t-1}_{-i}. Then, player i samples an action from her strategy $p^t_i \in \Delta(\mathcal{A})$ such that:

$$p^t_i(a_i) = \frac{e^{\beta U_i(a_i, a^{t-1}_{-i})}}{\sum_{a'_i\in\mathcal{A}} e^{\beta U_i(a'_i, a^{t-1}_{-i})}}, \quad \forall a_i \in \mathcal{A}, \qquad (3)$$

where parameter β measures a player's rationality: for large β player i is more likely to select a best response $a^t_i \in \arg\max_{a_i\in\mathcal{A}} U_i(a_i, a^{t-1}_{-i})$; and for $\beta = 0$ player i samples a^t_i uniformly. Moreover, the strategy p^t_i is myopic as it depends only on the other players' last actions a^{t-1}_{-i}.

[1] P^* satisfies $\mu(a)P^*(a,\tilde{a}) = \mu(\tilde{a})P(\tilde{a},a)\forall a,\tilde{a} \in \mathcal{A}^N$. The chain is called time-reversible if $P^* = P$.

Log-linear learning induces an irreducible and aperiodic Markov chain $\{X_t\}_{t\in\mathbb{Z}_+}$ with a time-reversible transition matrix $P \in \mathbb{R}^{\mathcal{A}\times\mathcal{A}}$ (Marden and Shamma, 2012) given by:

$$P_{a,\tilde{a}} = \frac{1}{N}\frac{e^{\beta U_i(\tilde{a}_i,\tilde{a}_{-i})}}{\sum_{a_i'\in\mathcal{A}_i} e^{\beta U_i(a_i',\tilde{a}_{-i})}}\mathbb{1}_{\tilde{a}\in\mathcal{N}(a)}, \tag{4}$$

where $\mathcal{N}(a) = \{\tilde{a} \in \mathcal{A}^N \mid \exists i \in [N] : \tilde{a}_{-i} = a_{-i}\}$. The stationary distribution $\mu \in \Delta(\mathcal{A}^N)$ of log-linear learning is given by (Blume, 1993):

$$\mu(a) = \frac{e^{\beta\Phi(a)}}{\sum_{\tilde{a}\in A^N} e^{\beta\Phi(\tilde{a})}} \quad \forall a \in \mathcal{A}^N. \tag{5}$$

The above can be verified by checking the detailed balance equations given by $\mu(a)P_{a,\tilde{a}} = \mu(\tilde{a})P_{\tilde{a},a}$ for all $a, \tilde{a} \in A^N$. For $\beta \to \infty$, sampling an action profile $a \in \mathcal{A}^N$ from the stationary distribution μ returns a maximizer of the potential $\Phi(\cdot)$ with arbitrarily high probability. Thus, when all players adhere to log-linear learning with sufficiently large β, the global outcome, in the long run, will correspond to a potential maximizer, *i.e.,* an efficient Nash equilibrium.

For sufficiently large β, it was shown that log-linear learning converges asymptotically to a potential maximizer and thus an efficient Nash equilibrium (Blume, 1993; Young, 1993; Marden and Shamma, 2012). Except for a few works (Montanari and Saberi, 2010; Asadpour and Saberi, 2009; Shah and Shin, 2010), none of the previous works, however, have finite-time convergence guarantees to such an efficient Nash equilibrium, and (Montanari and Saberi, 2010; Asadpour and Saberi, 2009; Shah and Shin, 2010) have additional assumptions on the potential game. Thus, in the following section, we establish a bound on the convergence time of log-linear learning in general potential games.

3.2. Convergence time of log-linear learning

We now state our main result on the convergence time of log-linear learning to an ϵ-efficient NE. Before stating the result we briefly introduce some notation. Denote by a^* a potential maximizer, *i.e.,* $a^* \in \arg\max_{a\in\mathcal{A}^N} \Phi(a)$. The set of ϵ-optimal action profiles is defined as $\mathcal{A}^N(\epsilon) := \{a \in \mathcal{A}^N | \Phi(a) \geq \Phi(a^*) - \epsilon\}$ with cardinality $A^N(\epsilon) = |\mathcal{A}^N(\epsilon)|$. Furthermore, the suboptimality gap is defined as $\Delta := \min_{a\in\mathcal{A}^N:\Phi(a)<\Phi(a^*)} (\Phi(a^*) - \Phi(a))$ and is non-negative.

Theorem 3.1. *Consider a potential game with a potential function $\Phi : \mathcal{A}^N \to [0, 1]$ with $A \geq 4$.*[2] *For any $\epsilon \in (0, 1)$ and any initial distribution μ^0, assume that players adhere to log-linear learning with:*

$$\beta \geq \frac{1}{\max\{\epsilon/2, \Delta\}} \log\left((A^N - A^N(\epsilon/2))\left(\frac{4}{\epsilon A^N(\epsilon/2)} - \frac{1}{A^N(\epsilon/2)}\right)\right). \tag{6}$$

Then,

$$\mathbb{E}_{a\sim\mu^t}[\Phi(a)] \geq \max_{a\in\mathcal{A}^N} \Phi(a) - \epsilon,$$

for

$$t \geq \frac{25N^2A^5}{16\pi^2}e^{4\beta}\left(\log\log A^N + \log\beta + 2\log\frac{4}{\epsilon}\right).$$

In other words, after $t = \tilde{\Omega}(N^2A^5(\frac{A^N}{\epsilon})^{1/\max\{\epsilon,\Delta\}})$ rounds of log-linear learning with $\beta = \Omega(\frac{1}{\max\{\epsilon,\Delta\}}\log\frac{A^N}{\epsilon})$ the expected value of the potential of the joint action at time t is ϵ-optimal.

Then,

$$\mathbb{E}_{a\sim\mu^t}[\Phi(a)] \geq \max_{a\in\mathcal{A}^N} \Phi(a) - \epsilon,$$

for

$$t \geq \frac{25N^2A^5}{16\pi^2}e^{4\beta}\left(\log\log A^N + \log\beta + 2\log\frac{2}{\epsilon}\right).$$

In other words, after $t = \tilde{\Omega}(N^2A^5(\frac{A^N}{\epsilon})^{1/\max\{\epsilon,\Delta\}})$ rounds of log-linear learning with $\beta = \Omega(\frac{1}{\max\{\epsilon,\Delta\}}\log\frac{A^N}{\epsilon})$ the expected value of the potential of the joint action at time t is ϵ-optimal.

Theorem 3.1 provides the first finite-time convergence rates to an ϵ-efficient NE in general potential games. In order to be ϵ-close to an efficient NE β must scale as $\Omega(\frac{1}{\max\{\epsilon,\Delta\}}\log\frac{A^N}{\epsilon})$. Furthermore, since the convergence time scales as $e^{4\beta}$ it grows polynomially in $\mathcal{A}$ and $1/\epsilon$ and exponentially in N. The exponential dependence on N is unavoidable without further assumptions on the game. However, by deriving problem-dependent bounds we are the first to avoid the exponential dependence in $1/\epsilon$, see Table 1. Note that the case $\Delta = 0$ is trivial as all action profiles are efficient NEs.

Proof outline. Here, we provide an outline of the proof, the full proof is deferred to Appendix B.1. The argument begins with the following decomposition:

$$\mathbb{E}_{a\sim\mu^t}[\Phi(a)] \geq \underbrace{\mathbb{E}_{a\sim\mu}[\Phi(a)]}_{\text{First term}} - 2\underbrace{\|\mu^t - \mu\|_{TV}}_{\text{Second term}}\underbrace{\max_{a\in\mathcal{A}^N}\Phi(a)}_{\leq 1}$$

[2]The assumption $A \geq 4$ is needed to lower-bound the log-Sobolev constant, see Lemma B.1.

Table 1. Convergence time of log-linear learning to an ϵ-efficient NE.

Works	Game setting	Assumptions	Convergence time
(Asadpour and Saberi, 2009)	Routing game with K vertices	Cost functions of degree at most p	$\tilde{\Omega}(e^{\frac{N}{\epsilon}})$
(Shah and Shin, 2010)	Potential game with interchangeable players	λ-Lipschitz continuous potential function	$\tilde{\Omega}(N(\frac{A\lambda}{\epsilon})^{\frac{A}{\epsilon}})$
Corollary B.3	Potential game with interchangeable players	$A \geq 4$	$\tilde{\Omega}(N(\frac{N^A}{\epsilon})^{\frac{1}{\max\{\epsilon,\Delta\}}})$
Theorem 3.1	Potential game	$A \geq 4$	$\tilde{\Omega}(N^2A^5(\frac{A^N}{\epsilon})^{\frac{1}{\max\{\epsilon,\Delta\}}})$

First term: To control the first term above, we rely on the novel lemma below which shows that an action profile sampled from μ is in expectation $\epsilon/2$-optimal if β is sufficiently large.

Lemma 3.2. *For any $\epsilon \in (0,1)$, if all players adhere to log-linear learning with:*

$$\beta \geq \frac{1}{\max\{\epsilon,\Delta\}} \log\left((A^N - A^N(\epsilon))\left(\frac{1}{\epsilon A^N(\epsilon)} - \frac{1}{A^N(\epsilon)}\right)\right),$$

then it holds that $\mathbb{E}_{a\sim\mu}[\Phi(a)] \geq \max_{a\in\mathcal{A}^N}\Phi(a) - \epsilon$.

The proof of this lemma is provided in Appendix B.2. We use it to control the first term as:

$$\mathbb{E}_{a\sim\mu}[\Phi(a)] \geq \max_{a\in\mathcal{A}^N}\Phi(a) - \epsilon/2$$

when β is defined as in Equation (6).

Second term: We can control the mixing time of log-linear learning using the bound of Lemma 2.2:

$$\|\mu^t - \mu\|_{TV} \leq \epsilon/4$$

for $t \geq \frac{1}{\rho(PP^*)}(\log\log\frac{1}{\mu_{\min}} + 2\log\frac{4}{\epsilon})$. We then bound $\rho(PP^*)$ using the novel lemma below.

Lemma 3.3. *Consider a Markov chain over state space $\mathcal{A}^N$ with $A \geq 4$. Assume that there exists $p_{\min}, p_{\max} \in (0,1]$, such that the corresponding transition matrix $P \in \mathbb{R}^{A^N\times A^N}$ satisfies:*

$$\frac{1}{N}p_{\min}\mathbb{1}_{\tilde{a}\in\mathcal{N}(a)} \leq P_{a,\tilde{a}} \leq \min\{1, \frac{1}{N}p_{\max}\}\mathbb{1}_{\tilde{a}\in\mathcal{N}(a)} \quad (7)$$

where $\mathcal{N}(a) = \{\tilde{a} \in \mathcal{A}^N \mid \exists i \in [N] : \tilde{a}_{-i} = a_{-i}\}$. Then the log-Sobolev constant $\rho(PP^*)$ of matrix PP^* is lower bounded by:

$$\rho(PP^*) \geq \frac{16\pi^2 A^{N-2}\mu_{\min}p_{\min}^3}{25N^2},$$

where μ is the stationary distribution of the Markov chain induced by P and $\mu_{\min} = \min_{a\in\mathcal{A}^N}\mu(a)$.

The proof of this lemma is provided in Appendix B.3. It provides a bound on $\rho(PP^*)$ for any transition matrix that satisfies Equation (7). It is in particular applicable to the Markov chain induced by log-linear learning since the transition matrix specified in Equation (4) satisfies Equation (7).

Combination: Combining the two parts of the proof we deduce that:

$$\begin{aligned}\mathbb{E}_{a\sim\mu^t}[\Phi(a)] &\geq \mathbb{E}_{a\sim\mu}[\Phi(a)] - 2\|\mu^t - \mu\|_{TV}\max_{a\in\mathcal{A}^N}\Phi(a) \\ &\geq \max_{a\in\mathcal{A}^N}\Phi(a) - \frac{\epsilon}{2} - \frac{2\epsilon}{4},\end{aligned}$$

for $t \geq \frac{25N^2A^5}{16\pi^2}e^{4\beta}\left(\log\log A^N + \log\beta + 2\log\frac{4}{\epsilon}\right)$. This concludes the proof of the theorem. □

In the following, we additionally assume that the potential game is symmetric, *i.e.,* players are interchangeable.[3] In this setting, we derive a bound on the convergence time of log-linear learning that depends polynomially not only on $1/\epsilon$ but also on N.

Corollary 3.4 (Sketch). *Consider a symmetric potential game. For any $\epsilon \in (0,1)$ and initial distribution μ^0, assume that players adhere to modified log-linear learning (Shah and Shin, 2010) with $\beta = \Omega(\frac{1}{\max\{\epsilon,\Delta\}}\log(\frac{N^A}{\epsilon}))$. Then, $\mathbb{E}_{a\sim\mu^t}[\Phi(a)] \geq \max_{a\in\mathcal{A}^N}\Phi(a) - \epsilon$ for $t = \tilde{\Omega}(N(\frac{N^A}{\epsilon})^{\frac{1}{\max\{\epsilon,\Delta\}}})$.*

We provide a full statement of the corollary and its proof in Appendix B.4. For symmetric potential games with a λ-Lipschitz-continuous potential function, Shah and Shin (2010) proves a convergence time of $\tilde{\Omega}(N(\frac{A\lambda}{\epsilon})^{\frac{A}{\epsilon}})$. In comparison, relaxing the Lipschitzness assumption comes at a small cost of polynomial dependence on N rather than a linear one. However, our result greatly improves the dependence on ϵ to a polynomial one.

[3] A definition of a symmetric game is given in Appendix B.4, see Definition B.2.

4. Relaxed structural assumption

In this section, we investigate how our convergence time results are affected when structural assumptions are relaxed. In particular, we analyze binary log-linear learning which handles reduced feedback, then we study perturbations of log-linear learning such as noisy utility observations.

4.1. Reduced feedback

Log-linear learning requires players to access their utilities for all possible actions given the other players' actions. Having such full-information feedback when action sets are large can be demanding. Binary log-linear learning (Arslan et al., 2007; Marden et al., 2007) alleviates this limitation by requiring two-point feedback, reducing the feedback needed by a factor A per round. Now, we briefly review the binary-log-linear learning rule. Then, we derive the first finite-time convergence bound of binary log-linear learning to an ϵ-efficient Nash equilibrium, showing that the deterioration in convergence time is a constant.

Binary log-linear learning proceeds as log-linear learning with the distinction that the player i allowed to alter her action first samples a trial action $\tilde{a}_i$ uniformly from her action set $\mathcal{A}$. She then plays according to the strategy:

$$p_i^t(a) = \begin{cases} \frac{e^{\beta U_i(a, a_{-i}^{t-1})}}{e^{\beta U_i(a_i^{t-1}, a_{-i}^{t-1})} + e^{\beta U_i(\tilde{a}_i, a_{-i}^{t-1})}} & \text{for } a \in \{a_i^{t-1}, \tilde{a}_i\} \\ 0 & \text{otherwise} \end{cases}$$

Here, player i can either repeat her action a_i^{t-1} or play one other randomly sampled action $\tilde{a}_i$ rather than any action $a_i \in \mathcal{A}$ as in log-linear learning.

Theorem 4.1. *Consider a potential game with potential function $\Phi : \mathcal{A}^N \to [0,1]$ and $A \geq 4$. For any $\epsilon \in (0,1)$ and initial distribution μ^0, assume that players adhere to binary log-linear learning with $\beta = \Omega(\frac{1}{\max\{\epsilon,\Delta\}} \log \frac{A^N}{\epsilon})$. Then, it holds that*

$$\mathbb{E}_{a\sim\mu^t}[\Phi(a)] \geq \max_{a\in\mathcal{A}^N} \Phi(a) - \epsilon,$$

for

$$t \geq \underbrace{\frac{25N^2A^5}{2\pi^2} e^{4\beta} \left(\log\log A^N + \log\beta + 2\log\frac{4}{\epsilon}\right)}_{\tilde{\Omega}\left(N^2A^5\left(\frac{A^N}{\epsilon}\right)^{\frac{N}{\max\{\epsilon,\Delta\}}}\right)}.$$

Theorem 4.1 shows the first finite-time convergence guarantee for binary log-linear learning, we provide its detailed proof in Appendix C. It is remarkable that with significantly less feedback per round, binary log-linear achieves the same convergence speed as log-linear learning up to a factor of 8 (Theorem 3.1). This raises the question of whether two-point feedback is sufficient for learning ϵ-efficient Nash equilibria or whether the convergence time bounds we proved for log-linear learning are loose.

4.2. Perturbed log-linear learning

Classical log-linear learning relies on two limiting assumptions: 1) Players have access to their exact utilities. However, in real-world applications, the presence of noise is typical as uncertainties and hidden factors generate inexact measurements. 2) Players are rational. However, empirical evidence suggests that players have limited rationality and therefore may occasionally deviate from the log-linear learning rule in practical scenarios. Our next result generalizes Theorem 3.1 to the case where the log-linear learning rule is subject to small perturbations. As we will show, this generalization can address the two limitations above.

We first derive a general statement for Markov chains which shows that the induced stationary distribution is Lipschitz-continuous as a function of the transition matrix. We then use this lemma to prove our main result on the convergence time of perturbed log-linear learning to an ϵ-efficient NE.

Lemma 4.2 (Lipschitzness)**.** *Consider two irreducible and aperiodic transition matrices $P_1, P_2 \in \mathbb{R}^{A^N\times A^N}$. Let μ_1 and μ_2 be the stationary distributions of the Markov chains induced by P_1 and P_2, respectively. Then, the following holds:*

$$\|\mu_1 - \mu_2\|_2 \leq \min\{L(P_1), L(P_2)\}\|P_1 - P_2\|_2,$$

where $L(P_k) := \frac{2A^N}{\rho(P_kP_k^*)}(\log\log\frac{1}{\mu_{k,\min}} + \log(8A^N))$ *and* $\mu_{k,\min} = \min_{a\in\mathcal{A}^N}\mu_k(a)$ *for* $k = 1, 2$.

We provide a proof in Appendix D.4. Compared to the result of (Zhang et al., 2023), we considerably improve the Lipschitz constant by using the mixing time bound based on log-Sobolev inequalities (Lemmas 2.2 and 3.3). In particular, (Zhang et al., 2023, Lemma 24) entails a Lipschitz constant $L = \tilde{\mathcal{O}}((e/p_{\min})^N)$ while Lemma 4.2 implies that $L = \tilde{\mathcal{O}}(1/(\mu_{\min}p_{\min}^3))$.[4] Now, we state the main result of this section.

Theorem 4.3. *Consider a potential game with a potential function $\Phi : \mathcal{A}^N \to [0,1]$ and $A \geq 4$. Furthermore, consider a learning rule with transition matrix P and assume that there exists $p_{\min}, p_{\max} \in (0,1]$, such that for all $a, \tilde{a} \in \mathcal{A}$ it holds that:*

$$\frac{1}{N}p_{\min}\mathbb{1}_{\tilde{a}\in\mathcal{N}(a)} \leq P_{a,\tilde{a}} \leq \min\{1, \frac{1}{N}p_{\max}\}\mathbb{1}_{\tilde{a}\in\mathcal{N}(a)}. \tag{8}$$

[4] This can be seen by injecting the log-Sobolev bound of lemma 3.3 into our Lemma above.

For any $\epsilon \in (0,1)$ and any initial distribution μ^0, assume all players adhere to this learning rule with $\beta = \Omega\left(\frac{1}{\max\{\epsilon,\Delta\}}\log\frac{1}{\epsilon}\right)$. Then,

$$\mathbb{E}_{a\sim\mu^t}[\Phi(a)] \geq \max_{a\in\mathcal{A}}\Phi(a) - \epsilon - L\sqrt{A^N}\|P - P_\ell\|_2,$$

for

$$t \geq \frac{25N^{3/2}e^N}{(2\pi)^{5/2}A^N p_{\min}^{N+3}}\log\left(\frac{4A^N}{\epsilon^2}\log\frac{e^N}{p_{\min}^N\sqrt{2\pi N}}\right),$$

where P_ℓ is the transition matrix of log-linear learning and L is a Lipschitz constant of order $\tilde{\mathcal{O}}\left(N^2A^{N+5}e^{\log(A^N/\epsilon)/\max\{\epsilon,\Delta\}}\right)$.

Theorem 4.3 shows that small perturbations of the log-linear learning rule do not compromise the convergence to an ϵ-efficient NE. In particular, if the players follow a learning rule P with $\|P - P_\ell\|_2 = \mathcal{O}(\epsilon/(L\sqrt{A^N}))$ then they converge to an ϵ-efficient NE in time polynomial in $1/\epsilon$. On the other hand, due to the unavailability of the stationary distribution of the perturbed learning rule, we suffer an extra factor of $(N/p_{\min})^N/N!$ in the convergence time guarantee compared to log-linear learning.

We now consider two explicit types of perturbations: Noisy utilities and a modified learning rule.

4.2.1. Corrupted utilities with additive noise

In the following, we assume that players observe noise-corrupted utilities $(\hat{U}_i)_{i\in[N]}$ which satisfy:

$$\hat{U}_i(a_i,a_{-i}) = U_i(a_i,a_{-i}) + \xi_i(a_i,a_{-i}), \quad \forall (a_i,a_{-i}) \in \mathcal{A}^N \tag{9}$$

where $\xi_i(a_i,a_{-i}) \in [-\xi,\xi]$ is a bounded noise term. Alternatively, the noise could be assumed to be centered i.i.d. random variables with bounded variance (Leslie and Marden, 2011).

Using Theorem 4.3, we hereafter prove that log-linear learning is robust to noisy feedback.

Corollary 4.4. *Consider the setting of Theorem 4.3 with noise-corrupted utilities as in Equation (9). If all players adhere to log-linear learning with $\beta = \Omega\left(\frac{1}{\max\{\epsilon,\Delta\}}\log\frac{1}{\epsilon}\right)$ and $\xi \leq 1/(2\beta)$, then*

$$\mathbb{E}_{a\sim\mu^t}[\Phi(a)] \geq \max_{a\in\mathcal{A}}\Phi(a) - \epsilon - \frac{7LA^{3N/2}}{2N}\beta\xi,$$

for $t = \Omega\left(N^{3/2}A^3e^{N+\beta(1+2\xi)(N+3)}\log\frac{1}{\epsilon^2}\right)$ with $L = \tilde{\mathcal{O}}\left(N^2A^{N+5}e^{\log(A^N/\epsilon)/\max\{\epsilon,\Delta\}}\right)$.

The proof is provided in Appendix D.2. Corollary 4.4 shows that log-linear learning with corrupted utilities converges to an ϵ-efficient NE in time polynomial in $1/\epsilon$ if the corruption magnitude ξ is sufficiently small. Our finite-time convergence result extends previous works on robust learning which provide asymptotic guarantees (Leslie and Marden, 2011; Lim and Shamma, 2013; Bravo and Mertikopoulos, 2017). The key to this result lies in showing that the transition matrix of the Markov chain induced by corrupted utilities is close to its corruption-free counterpart.

4.2.2. Log-linear learning mixed with uniform exploration

In the following, we assume players occasionally explore actions randomly. A modification of log-linear learning based on the fixed-share algorithm (Herbster and Warmuth, 1998) can reflect such a random behavior. In the so-called fixed-share log-linear learning, a player i is randomly chosen and allowed to alter her action. Player i samples her new action from the following distribution:

$$\hat{p}_i^t(a_i) = \frac{\xi}{A} + \frac{(1-\xi)e^{\beta U_i(a_i,a_{-i}^{t-1})}}{\sum_{a_i'\in\mathcal{A}}e^{\beta U_i(a_i',a_{-i}^{t-1})}}, \quad \forall a_i \in \mathcal{A}.$$

The exploration parameter $\xi \in (0,1)$ determines how likely a player is to act randomly, where a value of $\xi = 1$ corresponds to a uniform action sampling while $\xi = 0$ corresponds to standard log-linear learning. For simplicity, we focus on the full-information case but fixed-share log-linear learning can easily be adapted to the binary setting. Note that this modification resembles the ϵ-Hedge strategy (Heliou et al., 2017) in the expert advice literature, and under binary feedback, this modification resembles the Epx3.P strategy (Auer et al., 2002; Bubeck et al., 2012) in the multi-armed bandit literature. Here, the fixed share ξ/A ensures a lower bound on the exploration.

Without explicit knowledge of the stationary distribution of this learning rule, we can apply Theorem 4.3 to deduce the following result.

Corollary 4.5. *Consider the setting of Theorem 4.3, where all players adhere to fixed-share log-linear learning with $\beta = \Omega\left(\frac{1}{\max\{\epsilon,\Delta\}}\log\frac{1}{\epsilon}\right)$. Then, for any $\epsilon \in (0,1)$ and initial distribution μ^0 we have:*

$$\mathbb{E}_{a\sim\mu^t}[\Phi(a)] \geq \max_{a\in\mathcal{A}}\Phi(a) - \epsilon - \frac{LA^N}{\sqrt{N}}\xi,$$

for $t = \Omega\left(N^{3/2}A^{N+3}e^{\beta(N+3)}/(1-\xi)^{N+3}\right)$ with $L = \tilde{\mathcal{O}}\left(N^2A^{N+5}e^{\log(A^N/\epsilon)/\max\{\epsilon,\Delta\}}\right)$.

We provide a proof in Appendix D.3. Corollary 4.5 guarantees the convergence of fixed-share log-linear learning to an ϵ-efficient Nash equilibrium in time polynomial in $1/\epsilon$ if the exploration parameter ξ is sufficiently small. The key to this result is to show that the transition matrix of fixed-share log-linear learning is close to the transition matrix of the unperturbed learning rule in terms of the ℓ_2 distance.

5. Conclusion

We provided the first finite-time convergence guarantees to an ϵ-efficient NE for potential games using a novel mixing-time bound based on a log-Sobolev constant. In particular, we guarantee a polynomial dependence on $1/\epsilon$ using a problem-dependent analysis. Furthermore, under the additional assumption that the game is symmetric, we showed that the exponential dependence on the number of players N present in our bound on the convergence time can be avoided. To deal with reduced feedback, *i.e.,* two-point feedback on the utility, we considered binary log-linear learning and showed that it enjoys the same convergence time as log-linear learning up to numerical constants. Finally, we proved that the convergence time of log-linear is not hindered by corruptions of the observed utilities by bounded noise or by small perturbations in the learning rule. The relevance of this result is twofold: First, our analysis does not rely on characterizing the stationary distribution of this perturbed Markov chain; Second, the presence of noise is ubiquitous in real-world applications and it is therefore crucial that the implemented learning rule is robust to such corruptions.

In future work, we are interested in providing lower bounds on the convergence time of log-linear learning to an ϵ-efficient NE. Such bounds would allow us to assess the tightness of our results.

ACKNOWLEDGMENT

This work was gratefully supported by Swiss National Science Foundation, under the NCCR Automation Grant.

References

Gürdal Arslan, Jason R Marden, and Jeff S Shamma. Autonomous vehicle-target assignment: A game-theoretical formulation. 2007.

Arash Asadpour and Amin Saberi. On the inefficiency ratio of stable equilibria in congestion games. In *Internet and Network Economics: 5th International Workshop, WINE 2009, Rome, Italy, December 14-18, 2009. Proceedings 5*. Springer, 2009.

Peter Auer, Nicolo Cesa-Bianchi, Yoav Freund, and Robert E Schapire. The nonstochastic multiarmed bandit problem. *SIAM journal on computing*, 2002.

Baruch Awerbuch, Yossi Azar, Amir Epstein, Vahab Seyed Mirrokni, and Alexander Skopalik. Fast convergence to nearly optimal solutions in potential games. In *Proceedings of the 9th ACM conference on Electronic commerce*, 2008.

Lawrence E Blume. The statistical mechanics of strategic interaction. *Games and economic behavior*, 1993.

Mario Bravo and Panayotis Mertikopoulos. On the robustness of learning in games with stochastically perturbed payoff observations . *Games and Economic Behavior*, 2017.

Sébastien Bubeck, Nicolo Cesa-Bianchi, et al. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends® in Machine Learning*, 2012.

Samuel Burer and Adam N Letchford. Non-convex mixed-integer nonlinear programming: A survey. *Surveys in Operations Research and Management Science*, 2012.

Steve Chien and Alistair Sinclair. Convergence to approximate Nash equilibria in congestion games. *Games and Economic Behavior*, 2011.

Persi Diaconis and Laurent Saloff-Coste. Logarithmic Sobolev inequalities for finite Markov chains. *The Annals of Applied Probability*, 1996.

Amélie Heliou, Johanne Cohen, and Panayotis Mertikopoulos. Learning with bandit feedback in potential games. *Advances in Neural Information Processing Systems*, 2017.

Mark Herbster and Manfred K Warmuth. Tracking the best expert. *Machine learning*, 1998.

Walid Krichene, Benjamin Drighès, and Alexandre M Bayen. Online learning of nash equilibria in congestion games. *SIAM Journal on Control and Optimization*, 2015.

David S Leslie and Jason R Marden. Equilibrium selection in potential games with noisy rewards. In *International Conference on NETwork Games, Control and Optimization (NetGCooP 2011)*, 2011.

David A Levin and Yuval Peres. *Markov chains and mixing times*. American Mathematical Soc., 2017.

Yusun Lim and Jeff S Shamma. Robustness of stochastic stability in game theoretic learning. In *2013 American Control Conference*, 2013.

Jason R Marden and Jeff S Shamma. Revisiting log-linear learning: Asynchrony, completeness and payoff-based implementation . *Games and Economic Behavior*, 2012.

Jason R Marden and Jeff S Shamma. Game theory and distributed control. In *Handbook of game theory with economic applications*. Elsevier, 2015.

Jason R Marden and Adam Wierman. Distributed welfare games. *Operations Research*, 2013.

Jason R Marden, Gürdal Arslan, and Jeff S Shamma. Connections between cooperative control and potential games illustrated on the consensus problem . In *2007 European Control Conference (ECC)*. IEEE, 2007.

Laurent Miclo. *Remarques sur l'hypercontractivité et l'évolution de l'entropie pour des chaines de Markov finies* . Springer, 1997.

Dov Monderer and Lloyd S Shapley. Fictitious play property for games with identical interests. *Journal of economic theory*, 1996a.

Dov Monderer and Lloyd S Shapley. Potential games. *Games and economic behavior*, 1996b.

Andrea Montanari and Amin Saberi. Convergence to equilibrium in local interaction games and ising models. *arXiv preprint arXiv:0812.0198*, 2008.

Andrea Montanari and Amin Saberi. The spread of innovations in social networks. *Proceedings of the National Academy of Sciences*, 2010.

Ravi Montenegro, Prasad Tetali, et al. Mathematical aspects of mixing times in Markov chains. *Foundations and Trends® in Theoretical Computer Science*, 2006.

John Nash. Non-cooperative games. *Annals of mathematics*, 1951.

John F Nash Jr. Equilibrium points in n-person games. *Proceedings of the national academy of sciences*, 1950.

Dario Paccagnan, Rahul Chandan, and Jason R Marden. Utility and mechanism design in multi-agent systems: An overview. *Annual Reviews in Control*, 2022.

Gerasimos Palaiopanos, Ioannis Panageas, and Georgios Piliouras. Multiplicative weights update with constant step-size in congestion games: Convergence, limit cycles and chaos . *Advances in Neural Information Processing Systems*, 2017.

Robert W Rosenthal. A class of games possessing pure-strategy Nash equilibria. *International Journal of Game Theory*, 1973.

Devavrat Shah and Jinwoo Shin. Dynamics in congestion games. *ACM SIGMETRICS Performance Evaluation Review*, 2010.

H Peyton Young. The evolution of conventions. *Econometrica: Journal of the Econometric Society*, 1993.

Yizhou Zhang, Guannan Qu, Pan Xu, Yiheng Lin, Zaiwei Chen, and Adam Wierman. Global convergence of localized policy iteration in networked multi-agent reinforcement learning . *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 2023.

A. Markov chains and mixing times

In the following, we define the log-Sobolev constant and the spectral gap of a Markov chain. We then discuss the implications of Lemma 2.2 and provide a proof of the lemma.

Definition A.1 ((Diaconis and Saloff-Coste, 1996)**.** Consider a Markov chain $\{X_t\}_{t\in\mathbb{N}}$ over state space $\mathcal{A}^N$ with transition matrix P and stationary distribution μ. The log-Sobolev constant $\rho(P)$ is defined as:

$$\rho(P) := \inf_{\mathcal{L}_\pi(f^2)\neq 0} \frac{\mathcal{E}_P(f,f)}{\mathcal{L}_\pi(f^2)}. \tag{10}$$

The spectral gap $\lambda(P)$ is defined as:

$$\lambda(P) := \inf_{\mathrm{Var}_\pi(f)\neq 0} \frac{\mathcal{E}_P(f,f)}{\mathrm{Var}_\pi(f)}. \tag{11}$$

For any $f : \mathcal{A}^N \to \mathbb{R}$, the Dirichlet form $\mathcal{E}_P(f,f)$ is defined by:

$$\mathcal{E}_P(f,f) = \langle f, (I-P)f\rangle_\pi = \frac{1}{2}\sum_{a,\tilde{a}\in A^N} (f(a)-f(\tilde{a}))^2 P_{a,\tilde{a}}\mu(a),$$

the entropy-like quantity $\mathcal{L}(f^2)$ is defined by:

$$\mathcal{L}(f^2) = \sum_{a\in A^N} f(a)^2 \log\frac{f(a)^2}{\|f\|_2^2}\mu(a),$$

and the variance $\mathrm{Var}_\pi(f)$ is defined by:

$$\mathrm{Var}_\pi(f) = \sum_{a,\tilde{a}\in A^N} (f(a)-f(\tilde{a}))^2 \mu(a)\mu(\tilde{a}).$$

As mentioned in Section 2.2, using the log-Sobolev constant $\rho(PP^*)$ can often significantly improve classical mixing time bounds based on the spectral gap $\lambda(PP^*)$. Such classical bounds are of the form (Montenegro et al., 2006):

$$t^P_{\mathrm{mix}}(\epsilon) \le \frac{C}{\lambda(PP^*)}\left(\log\frac{1}{\sqrt{\mu_{\min}}} + \log\frac{1}{\epsilon}\right),$$

where C is some constant. Diaconis and Saloff-Coste (1996) in Lemma 3.1 show that the log-Sobolev constant $\rho(PP^*)$ is upper-bounded by the spectral gap $\lambda(PP^*)$ as follows: $2\rho(PP^*) \le \lambda(PP^*)$. Thus, roughly speaking if

$$\log\log\frac{1}{\mu_{\min}} \le \log\frac{1}{\sqrt{\mu_{\min}}} \tag{12}$$

then the mixing time bound based on the log-Sobolev constant is an improvement upon the mixing time bound based on the spectral gap. To make this more concrete, consider for example a Markov chain on the d-dimensional hypercube $\mathcal{H} = \{-1,1\}^d$ with uniform stationary distribution. Then, $\mu_{\min} = 2^{-d}$ and clearly Equation (12) is satisfied in this example.

Proof of Lemma 2.2. Let the relative entropy be defined as $D(\mu^t : \mu) := \sum_{a\in\mathcal{A}^N} \mu^t(a)\log\frac{\mu^t(a)}{\mu(a)}$. Then, for a Markov chain $\{X_t\}_{t\in\mathbb{N}}$ with irreducible transition matrix P the relative entropy $D(\mu^t : \mu)$ decays at the following rate (Miclo, 1997):

$$D(\mu^t : \mu) \le (1-\rho(PP^*))^t D(\mu^0 : \mu). \tag{13}$$

Using Pinsker's inequality we have that:

$$\|\mu^t - \mu\|_{TV} \leq \sqrt{\frac{D(\mu^t : \mu)}{2}} \leq \sqrt{(1-\rho(PP^*))^t D(\mu^0 : \mu)}. \tag{14}$$

Note that $\rho(PP^*) < 1$ since $2\rho(PP^*) \leq \lambda(PP^*)$ by Lemma 3.1 in (Diaconis and Saloff-Coste, 1996) and for the spectral gap $\lambda(PP^*)$ it is known that $\lambda(PP^*) < 1$ (Levin and Peres, 2017). To ensure that $\|\mu^t - \mu\|_{TV} \leq \epsilon$, we derive the following lower bound on t:

$$\begin{aligned}
&\sqrt{(1-\rho(PP^*))^t D(\mu^0 : \mu)} \leq \epsilon \\
\Leftrightarrow\quad & t \underbrace{\log(1-\rho(PP^*))}_{<0} \leq \log\left(\frac{\epsilon^2}{D(\mu^0 : \mu)}\right) \\
\Leftrightarrow\quad & t \geq \frac{1}{\log(1-\rho(PP^*))} \log\left(\frac{\epsilon^2}{D(\mu^0 : \mu)}\right) \\
\Leftrightarrow\quad & t \geq -\frac{1}{\rho(PP^*)} \log\left(\frac{\epsilon^2}{D(\mu^0 : \mu)}\right) \\
\Leftrightarrow\quad & t \geq \frac{1}{\rho(PP^*)} \left(\log(D(\mu^0 : \mu) + 2\log\left(\frac{1}{\epsilon}\right)\right) \\
\Leftrightarrow\quad & t \geq \frac{1}{\rho(PP^*)} \left(\log\log\left(\frac{1}{\mu_{\min}}\right) + 2\log\left(\frac{1}{\epsilon}\right)\right),
\end{aligned}$$

where in line 4 we used that $\log(1+x) \leq x$ for $x > -1$ and then in line 6 we used that for any μ_0, $D(\mu^0 : \mu) \leq \log \frac{1}{\mu_{\min}}$ for $\mu_{\min} := \min_{a \in \mathcal{A}^N} \mu(a)$. Thus, we conclude that the mixing time is upper-bounded as follows:

$$t^P_{\text{mix}}(\epsilon) \leq \frac{1}{\rho(PP^*)} \left(\log\log\frac{1}{\mu_{\min}} + 2\log\frac{1}{\epsilon}\right). \tag{15}$$

□

B. Convergence of log-linear learning

We first prove Lemma 3.2 and then state a novel lemma which we will use to prove Lemma 3.3. Lastly, we provide a formal proof of the convergence time of log-linear learning in Theorem 3.1.

B.1. Proof of Theorem 3.1

We are now ready to proceed with the proof of Theorem 3.1.

Proof of Theorem 3.1. By Lemma 3.3, the log-Sobolev constant $\rho(PP^*)$ can be lower-bounded as:

$$\rho(PP^*) \geq \frac{16\pi^2 A^N \mu_{\min} p^3_{\min}}{25N^2A^2} \geq \frac{16\pi^2 e^{-4\beta}}{25N^2A^5}, \tag{16}$$

where we used that by definition of P in (4) and μ in (5) $\mu_{\min}$ and $p_{\min}$ can be lower-bounded as follows:

$$\begin{aligned}
&\mu_{\min} = \min_{a \in \mathcal{A}^N} \mu(a) \geq \frac{e^{-\beta}}{A^N} \\
&P_{a,\tilde{a}} \geq \frac{e^{-\beta}}{NA}, \quad \forall \tilde{a} \in \mathcal{A}^N(a) \;\Rightarrow\; p_{\min} = \frac{e^{-\beta}}{A}.
\end{aligned}$$

Equation (2) in Lemma 2.2 provides an upper bound on the mixing time:

$$t^P_{\text{mix}}(\epsilon/4) \leq \frac{1}{\rho(PP^*)} \left(\log\log\frac{1}{\mu_*} + 2\log\frac{4}{\epsilon}\right).$$

Plugging the bound on the log-Sobolev constant from Equation (16) into the equation above we obtain:

$$\begin{aligned} t^P_{\text{mix}}(\epsilon/2) &\leq \frac{25N^2A^5}{16\pi^2}e^{4\beta}\bigg(\log\log\frac{1}{\mu_{\min}} + 2\log\frac{4}{\epsilon}\bigg) \\ &\leq \frac{25N^2A^5}{16\pi^2}e^{4\beta}\bigg(\log\log\frac{A^N}{e^{-\beta}} + 2\log\frac{4}{\epsilon}\bigg) \\ &\leq \frac{25N^2A^5}{16\pi^2}e^{4\beta}\bigg(\log\log A^N + \log\beta + 2\log\frac{4}{\epsilon}\bigg). \end{aligned}$$

By setting $t \geq \frac{25N^2A^5}{16\pi^2}e^{4\beta}\bigg(\log\log A^N + \log\beta + 2\log\frac{4}{\epsilon}\bigg)$ and β as in Equation (6) we obtain:

$$\begin{aligned} \mathbb{E}_{a\sim\mu^t}[\Phi(a)] &\geq \mathbb{E}_{a\sim\mu}[\Phi(a)] - 2\|\mu^t - \mu\|_{TV}\max_{a\in\mathcal{A}^N}\Phi(a) \\ &\geq \max_{a\in\mathcal{A}^N}\Phi(a) - \frac{\epsilon}{2} - \frac{2\epsilon}{4} \\ &= \max_{a\in\mathcal{A}^N}\Phi(a) - \epsilon, \end{aligned}$$

where the third line follows from Lemma 3.2, the fact that $\|\mu^t - \mu\|_{TV} \leq \epsilon/2$, and the fact that $\Phi(\cdot) \in [0,1]$. This concludes the proof of Theorem 3.1. ☐

B.2. Proof of Lemma 3.2

Proof of Lemma 3.2. Define the set $\mathcal{A}^N_* = \{a^* \in \mathcal{A}^N \mid a^* \in \arg\max_{a\in\mathcal{A}^N}\Phi(a)\}$ as the set of potential maximizers with cardinality $A^N_* = |\mathcal{A}^N_*|$. Then, the expected value of the potential function $\Phi(\cdot)$ over the stationary distribution μ of log-linear learning in (5) can be bounded as follows:

$$\begin{aligned} \mathbb{E}_{a\sim\mu}[\Phi(a)] &= \sum_{a\in\mathcal{A}^N}\frac{e^{\beta\Phi(a)}}{\sum_{\tilde{a}\in\mathcal{A}^N}e^{\beta\Phi(\tilde{a})}}\Phi(a) \\ &\geq \sum_{a\in\mathcal{A}^N_*}\frac{e^{\beta\Phi(a)}}{\sum_{\tilde{a}\in\mathcal{A}^N}e^{\beta\Phi(\tilde{a})}}\Phi(a) \\ &= \sum_{a\in\mathcal{A}^N_*}\frac{\Phi(a)}{\sum_{\tilde{a}\in\mathcal{A}^N(\epsilon)}e^{\beta(\Phi(\tilde{a})-\Phi(a))} + \sum_{\tilde{a}\in\mathcal{A}^N\setminus\mathcal{A}^N(\epsilon)}e^{\beta(\Phi(\tilde{a})-\Phi(a))}} \\ &\geq \sum_{a\in\mathcal{A}^N_*}\frac{\Phi(a)}{\sum_{\tilde{a}\in\mathcal{A}^N(\epsilon)}e^{0} + \sum_{\tilde{a}\in\mathcal{A}^N\setminus\mathcal{A}^N(\epsilon)}e^{-\beta\min\{\epsilon,\Delta\}}} \\ &\geq \frac{A^N_*}{A^N(\epsilon) + (A^N - A^N(\epsilon))e^{-\beta\min\{\epsilon,\Delta\}}}\Phi(a^*), \end{aligned} \tag{17}$$

where the suboptimality gap Δ is given by $\Delta := \min_{a\in\mathcal{A}^N:\Phi(a)<\Phi(a^*)}(\Phi(a^*) - \Phi(a))$ with $a^* \in \mathcal{A}^N_*$. Then, we have that:

$$\begin{aligned} &\beta \geq \frac{1}{\max\{\epsilon,\Delta\}}\log\Bigg((A^N - A^N(\epsilon))\bigg(\frac{1}{\epsilon A^N(\epsilon)} - \frac{1}{A^N(\epsilon)}\bigg)\Bigg) \\ \implies\quad & e^{\beta\max\{\epsilon,\Delta\}} \geq (A^N - A^N(\epsilon))\frac{1-\epsilon}{\epsilon A^N(\epsilon)} \\ \implies\quad & \frac{A^N_* - A^N(\epsilon) + \epsilon A^N(\epsilon)}{1-\epsilon} \geq (A^N - A^N(\epsilon))e^{-\beta\min\{\epsilon,\Delta\}} \\ \implies\quad & \frac{A^N_* - (1-\epsilon)A^N(\epsilon)}{1-\epsilon} \geq (A^N - A^N(\epsilon))e^{-\beta\min\{\epsilon,\Delta\}} \\ \implies\quad & \frac{A^N_*}{A^N(\epsilon) + (A^N - A^N(\epsilon))e^{-\beta\min\{\epsilon,\Delta\}}} \geq 1-\epsilon. \end{aligned}$$

By injecting the last inequality into Equation (17) we deduce that for $\beta \geq \frac{1}{\max\{\epsilon,\Delta\}} \log(A^N - A^N(\epsilon))(\frac{1}{\epsilon A^N(\epsilon)} - \frac{1}{A^N(\epsilon)}))$, it holds that:

$$\mathbb{E}_{a\sim\mu}[\Phi(a)] \geq (1-\epsilon) \max_{a\in A^N} \Phi(a) \geq \max_{a\in A^N} \Phi(a) - \epsilon,$$

where we used that $\Phi(a) \leq 1$ for all $a \in \mathcal{A}^N$. □

B.3. Proof of Lemma 3.3

We first state the following lemma which will be used to prove Lemma 3.3.

Lemma B.1. *Consider the Markov chain $\{\hat{X}_t\}_{t\in\mathbb{N}}$ over a finite state space $\mathcal{A}^N$ with transition matrix $\hat{P} \in \mathbb{R}^{A^N \times A^N}$ given by:*

$$\hat{P}_{a,\tilde{a}} = \frac{1}{NA} \mathbb{1}_{\tilde{a}\in\mathcal{N}(a)} \tag{18}$$

where $\mathcal{N}(a) = \{\tilde{a} \in \mathcal{A}^N \mid \exists i \in [N] : \tilde{a}_{-i} = a_{-i}\}$. Assuming $A \geq 4$, the log-Soblev constant of $\hat{P}$ is lower-bounded by:

$$\rho(\hat{P}) \geq \frac{4\pi^2}{25NA}.$$

Proof. We will lower bound the log-Soblev constant of $\hat{P}$ in terms of the log-Soblev constant of another Markov chain for which a lower bound on the log-Soblev constant is known.

First, note that the Markov chain $\{\hat{X}_t\}_{t\in\mathbb{N}}$ is aperiodic and irreducible with stationary distribution $\hat{\mu}(a) = 1/A^N$. This can be verified by checking the detailed balance equations given by $\hat{\mu}(a)\hat{P}_{a,\tilde{a}} = \hat{\mu}(\tilde{a})\hat{P}_{\tilde{a},a}$ for all $a, \tilde{a} \in A^N$. Next, we make use of Corollary 2.15 in (Montenegro et al., 2006) to lower-bound $\rho(\hat{P})$ in terms of the log-Soblev constant of another Markov chain $\bar{X}_t$ with transition matrix $\bar{P}$ and stationary distribution $\bar{\mu}$. Namely,

$$\rho(\hat{P}) \geq \frac{1}{MC}\rho(\bar{P}),$$

where

$$M = \max_{a\in\mathcal{A}^N} \frac{\hat{\mu}(a)}{\bar{\mu}(a)},$$

$$C = \max_{a\neq\tilde{a}:\hat{P}_{a,\tilde{a}}\neq 0} \frac{\bar{\mu}(a)\bar{P}_{a,\tilde{a}}}{\hat{\mu}(a)\hat{P}_{a,\tilde{a}}}.$$

As the comparison Markov chain, we consider the product chain $\{\bar{X}_t\}_{t\in\mathbb{N}}$ with $\bar{X}_t = \prod_{i=1}^N \bar{X}_{i,t}$ on the state space $\mathbb{Z}_{K^N} = \prod_{i=1}^N \mathbb{Z}_K$. Here, each $\{\bar{X}_{i_t}\}_{t\in\mathbb{N}}$ is a simple random walk on the finite circle $\mathbb{Z}_K = \{1,\ldots,K\}$ with $K \geq 4$. The transition matrix and the stationary distribution of the simple random walk $\bar{X}_{i_t}$ are given by $\bar{P}_{i_{k,k\pm1}} = \frac{1}{2}$ and $\bar{\mu}_i(k) = 1/K$, respectively. Furthermore, the log-Soblev constant $\rho(\bar{P}_i)$ is lower bounded by $\rho(\bar{P}_i) \geq \frac{8\pi^2}{25K^2}$ (**?**)Example 4.2]diaconis1996logarithmic. Thus, by definition the product chain $\bar{X}_t$ (**?**)Sec. 2.5]diaconis1996logarithmic has the following transition matrix:

$$\bar{P}_{\mathbf{k},\tilde{\mathbf{k}}} = \frac{1}{2N}\mathbb{1}_{\tilde{\mathbf{k}}=(k_i\pm1,\mathbf{k}_{-i})},$$

and stationary distribution:

$$\bar{\mu}(\mathbf{k}) = \prod_{i=1}^N \bar{\mu}_i(k_i) = \prod_{i=1}^N \frac{1}{K} = \frac{1}{K^N}.$$

Furthermore, by Lemma 3.2 in (Diaconis and Saloff-Coste, 1996) the log-Soblev constant $\rho(\bar{P})$ of the product chain $\{\bar{X}_t\}_{t\in\mathbb{N}}$ is lower bounded by:

$$\rho(\bar{P}) = \frac{1}{N} \min_{i\in\{1,\ldots,N\}} \rho(\bar{P}_i) \geq \frac{8\pi^2}{25NK^2}. \tag{19}$$

Next, note that there is a one-to-one mapping between the set A^N and the set $\mathbb{Z}_K$ with $|\mathcal{A}| = A = K$ and thus a one-to-one mapping between the set $\mathcal{A}^N$ and the set $\mathbb{Z}_{K^N}$ with with $|\mathcal{A}^N| = A^N = K^N$. Therefore, we can assume that the Markov chains $\hat{X}_t$ and $\bar{X}_t$ operate on the same state space. By Equation (19) and with the following upper bounds on M and C:

$$M = \max_{a \in \mathcal{A}^N} \frac{\hat{\mu}(a)}{\bar{\mu}(a)} = \frac{A^N}{A^N} = 1$$
$$C = \max_{a \neq \tilde{a}: \hat{P}_{a,\tilde{a}} \neq 0} \frac{\bar{\mu}(a)\bar{P}_{a,\tilde{a}}}{\hat{\mu}(a)\hat{P}_{a,\tilde{a}}} = \frac{A}{2}.$$

the log-Soblev constant $\rho(\hat{P})$ can be lower-bounded by:

$$\rho(\hat{P}) \geq \frac{1}{MC}\rho(\bar{P}) \geq \frac{16\pi^2}{25NA^3},$$

which concludes the proof. □

Next, we proceed to prove Lemma 3.3.

Proof of Lemma 3.3. The Markov chain X_t with transition matrix P defined in (7) is aperiodic and irreducible and thus a unique stationary distribution μ exists with $\mu^t = \mu^0 P^t \to \mu$ for $t \to \infty$ from any initial distribution μ^0. We define $\mu_{\max}$ and $\mu_{\min}$ as $\max_{a \in \mathcal{A}} \mu(a) \leq \mu_{\max} \leq 1$ and $\min_{a \in \mathcal{A}} \mu(a) \geq \mu_{\min} > 0$, where $\mu_{\min} > 0$ follows from the irreducibility of X_t.

Next, we consider the modified Markov chain $\{X_t^*\}_{t \in \mathbb{N}}$ with transition matrix PP^* which is aperiodic and irreducible since X_t is aperiodic and irreducible. Concretely, since P contains self-loops, *i.e.,* $P_{a,a} > 0$, it follows that PP^* contains self-loops, *i.e.,*

$$PP^*_{a,a} = \sum_{a' \in \mathcal{A}} P_{a,a'} P^*_{a',a} = \sum_{a' \in \mathcal{A}} P_{a,a'} \frac{\mu(a) P_{a,a'}}{\mu(a')} \geq P_{a,a} P_{a,a} > 0,$$

and thus X_t^* is aperiodic. Furthermore, for any $a, \tilde{a} \in \mathcal{A}^N$:

$$\begin{aligned}
(PP^*)^N_{a,\tilde{a}} &= \sum_{\substack{a_l \in \mathcal{A}^N \\ l=1,\ldots,N-1}} (PP^*)_{a,a_1} \ldots (PP^*)_{a_{N-1},\tilde{a}} \\
&= \sum_{\substack{a_l \in \mathcal{A}^N \\ l=1,\ldots,N-1}} \sum_{a' \in \mathcal{A}^N} P_{a,a'} P^*_{a',a_1} \cdots \sum_{a' \in \mathcal{A}^N} P_{a_{N-1},a'} P^*_{a',\tilde{a}} \\
&\geq \sum_{\substack{a_l \in \mathcal{A}^N \\ l=1,\ldots,N-1}} P_{a,a_1} P_{a_1,a_1} \ldots P_{a_{N-1},\tilde{a}} P_{\tilde{a},\tilde{a}} > 0,
\end{aligned}$$

where we used that $P^N_{a,\tilde{a}} > 0$ and $P_{a,a} > 0$ for all $a, \tilde{a} \in \mathcal{A}^N$ as well as the identity $\mu(a) P^*_{a,\tilde{a}} = \mu(\tilde{a}) P_{\tilde{a},a}$. It follows that X_t^* is irreducible. Thus, for X_t^* a unique stationary distribution μ^* exists. By Proposition 1.23 in (Levin and Peres, 2017) the stationary distribution of P^* is given by μ and thus the stationary distribution of PP^* is given by μ since $\mu PP^* = \mu P^* = \mu$. Furthermore, the following holds for the transition matrix PP^* :

$$\frac{1}{N} p^2_{\min} \mathbb{1}_{\tilde{a} \in \mathcal{A}^N(a)} \leq (PP^*)_{a,\tilde{a}} \leq \mathbb{1}_{\tilde{a} \in \mathcal{A}^N(a)}$$

where

$$(PP^*)_{a,\tilde{a}} = \sum_{a' \in \mathcal{A}} P_{a,a'} P^*_{a',\tilde{a}} \geq P_{a,\tilde{a}} P^*_{\tilde{a},\tilde{a}} \geq P_{a,\tilde{a}} P^*_{\tilde{a},\tilde{a}} \geq \frac{p^2_{\min}}{N}.$$

Next, we compute M and C, defined bellow, of the Markov chains X_t^* and X_t:

$$M = \max_{a \in \mathcal{A}^N} \frac{\mu(a)}{\mu(a)} = 1$$
$$C = \max_{a \neq \tilde{a} : (PP^*)_{a,\tilde{a}} \neq 0} \frac{\mu(a) P_{a,\tilde{a}}}{\mu(a)(PP^*)_{a,\tilde{a}}} \leq \frac{N}{p_{\min}^2}.$$

From Corollary 2.15 in (Diaconis and Saloff-Coste, 1996), it follows that the log-Soblev constant $\rho(PP^*)$ is lower-bounded by:

$$\rho(PP^*) \geq \frac{1}{MC}\rho(P) \geq \frac{p_{\min}^2}{N}\rho(P). \tag{20}$$

Next, we compare the Markov chain X_t to the Markov chain $\hat{X}_t$ with transition matrix $\hat{P}$ specified in Equation (18) of Lemma B.1. To this end, we compute M and C of the Markov chains X_t and $\hat{X}_t$:

$$M = \max_{a \in \mathcal{A}^N} \frac{\mu(a)}{\hat{\mu}(a)} \leq A^N$$
$$C = \max_{a \neq \tilde{a} : P_{a,\tilde{a}} \neq 0} \frac{\hat{\mu}(a)\hat{P}_{a,\tilde{a}}}{\mu(a) P_{a,\tilde{a}}} \leq \frac{N}{A^N N A \mu_{\min} p_{\min}},$$

Thus, by Corollary 2.15 in (Diaconis and Saloff-Coste, 1996) and by Lemma B.1 the log-Soblev constant $\rho(P)$ can be lower-bounded by:

$$\rho(P) \geq \frac{1}{MC}\rho(\hat{P}) \geq A^N A \mu_{\min} p_{\min} \rho(\hat{P}) \geq \frac{16\pi^2 A^N \mu_{\min} p_{\min}}{25 N A^2}. \tag{21}$$

Combining Equation (20) and (21), we conclude that the log-Soblev constant $\rho(PP^*)$ is lower-bounded by:

$$\rho(PP^*) \geq \frac{16\pi^2 A^N \mu_{\min} p_{\min}^3}{25 N^2 A^2}.$$

□

B.4. Modified log-linear learning in symmetric potential games

In the following, we consider a symmetric potential game and show that in this setting, we obtain convergence time guarantees to an ϵ-efficient NE that depend polynomially on the number of players N and on $1/\epsilon$. This improves the convergence time result provided in (Shah and Shin, 2010) which shows an exponential dependence on $1/\epsilon$.

A game is said to be symmetric if it satisfies the following definition.

Definition B.2. A game is symmetric if for any permutation π of $\{1, \dots, N\}$ the following holds:

$$U_i(a_1, \dots, a_N) = U_{\pi(i)}(a_{\pi(1)}, \dots, a_{\pi(N)}).$$

In other words, the utility of player i depends only on how many players are playing each action $a \in \mathcal{A}$ and not on players' identities. Thus, in a symmetric potential game, if $A < N$, the potential function Φ can be redefined in terms of a lower-dimensional function $\Phi : \Psi^{\mathcal{A}} \to [0, 1]$, where

$$\Psi^{\mathcal{A}} := \left\{ \left(\frac{v_1}{N}, \dots, \frac{v_A}{N} \right) \mid v_j \in \mathbb{Z}_+ \; \forall j \in [A], \; \sum_{j=1}^{A} v_j = N \right\}$$

with cardinality $Y = |\Psi^{\mathcal{A}}| \leq (N+1)^{A-1}$. Then, for any $a = (a_1, \dots, a_N) \in \mathcal{A}^N$ we have $\Phi(a) = \Phi(x(a))$ with $x(a) = (x_1(a), \dots, x_N(a))$. Here $x_j(a)$ denotes the fraction of players that selected action $j \in A^N$, *i.e.,* $x_j(a) = 1/N|\{i \in [N] \mid a_i = j\}|$.

To obtain their results for symmetric potential games, Shah and Shin (2010) propose modified log-linear learning, a variant of log-linear learning. In modified log-linear learning every player i has an independent exponential clock of rate α/z_i^t, where $\alpha > 0$ is a parameter and $z_i^t := 1/N|\{j \in [N] \mid a_j^t = a_i^t\}|$.[5] This means that the times between two consecutive clock-ticks are independent and distributed as the exponential distribution of mean α/z_i^t. When the clock of player i ticks, she is allowed to alter her current action. Player i samples an action from her strategy $p_i^t \in \Delta(\mathcal{A})$ defined as in Equation (3). Modified log-linear learning induces an aperiodic and irreducible Markov chain on the lower-dimensional state space $\Psi^{\mathcal{A}}$ with stationary distribution $\mu_m \in \Delta(\Psi^{\mathcal{A}})$ given by (**?**)Lemma 2]shah2010dynamics:

$$\mu_m(x) = \frac{e^{\beta\Phi(x)}}{\sum_{\tilde{x}\in\Psi_N^{\mathcal{A}}} e^{\beta\Phi(\tilde{x})}} \quad \forall x \in \Psi^{\mathcal{A}}.$$

In the following, we show that in a symmetric potential game, if all players adhere to modified log-linear learning an ϵ-efficient NE is reached in time polynomial in N and $1/\epsilon$.

Corollary B.3. *Consider a symmetric potential game with potential function $\Phi : \Psi^{A^N} \to [0,1]$. For any $\epsilon \in (0,1)$ and any initial distribution μ_m^0, assume that players adhere to modified log-linear learning with:*

$$\beta = \Omega\left(\frac{1}{\max\{\epsilon,\Delta\}}\log\left(\frac{N^A}{\epsilon}\right)\right).$$

Then,

$$\mathbb{E}_{x\sim\mu_m^t}[\Phi(x)] \geq \max_{x\in\Psi^{\mathcal{A}}}\Phi(x) - \epsilon,$$

for

$$t \geq \frac{N}{\alpha c}e^{3\beta}\left(\log((A-1)\log(N+1)) + \log\beta + 2\log\frac{4}{\epsilon}\right) = \tilde{\Omega}\left(N\left(\frac{N^A}{\epsilon}\right)^{\frac{1}{\max\{\epsilon,\Delta\}}}\right),$$

where c is a constant that depends on A.

Proof. Define the set $\Psi_*^{\mathcal{A}} = \{x^* \in \Psi^{\mathcal{A}} \mid x^* \in \arg\max_{x\in\Psi^{\mathcal{A}}}\Phi(x)\}$ as the set of potential maximizers with cardinality $Y_* = |\Psi_*^{\mathcal{A}}|$ and define the set $\Psi^{\mathcal{A}}(\epsilon) = \{x \in \Psi^{\mathcal{A}} \mid \Phi(x) \geq \Phi(x^*)\}$ as the set of ϵ-approximate potential maximizers with cardinality $Y(\epsilon) = |\Psi^{\mathcal{A}}(\epsilon)|$. Then, the expected value of the potential function $\Phi(\cdot)$ over the stationary distribution μ_m of modified log-linear learning in (5) can be bounded as follows:

$$\begin{aligned}
\mathbb{E}_{x\sim\mu_m}[\Phi(x)] &= \sum_{a\in\Psi^{\mathcal{A}}} \frac{e^{\beta\Phi(x)}}{\sum_{\tilde{x}\in\Psi^{\mathcal{A}}} e^{\beta\Phi(\tilde{x})}}\Phi(x) \\
&= \sum_{a\in\Psi_*^{\mathcal{A}}} \frac{1}{\sum_{\tilde{x}\in\Psi^{\mathcal{A}}(\epsilon)} e^{\beta(\Phi(\tilde{x})-\Phi(x))} + \sum_{\tilde{x}\in\Psi^{\mathcal{A}}\setminus\Psi^{\mathcal{A}}(\epsilon)} e^{\beta(\Phi(\tilde{x})-\Phi(x))}}\Phi(x) \\
&\geq \sum_{a\in\Psi_*^{\mathcal{A}}} \frac{1}{\sum_{\tilde{x}\in\Psi^{\mathcal{A}}(\epsilon)} e^{0} + \sum_{\tilde{x}\in\Psi^{\mathcal{A}}\setminus\Psi^{\mathcal{A}}(\epsilon)} e^{-\beta\min\{\epsilon,\Delta\}}}\Phi(x) \\
&\geq \frac{Y_*}{Y(\epsilon) + (Y - Y(\epsilon))e^{-\beta\min\{\epsilon,\Delta\}}}\Phi(x^*),
\end{aligned}$$

where the suboptimality gap Δ is given by $\Delta := \min_{x\in\Psi^{\mathcal{A}}:\Phi(x)<\Phi(x^*)}(\Phi(x^*) - \Phi(x))$ with $x^* \in \Psi_*^{\mathcal{A}}$. Then, for

$$\beta \geq \frac{1}{\max\{\epsilon,\Delta\}}\log\left((N+1)^{A-1}\left(\frac{1}{\epsilon A^N(\epsilon)} - \frac{1}{A^N(\epsilon)}\right)\right), \tag{22}$$

it holds that

$$\frac{Y_*}{Y(\epsilon) + (Y - Y(\epsilon))e^{-\beta\min\{\epsilon,\Delta\}}} \geq 1 - \epsilon,$$

[5]The only change compared to log-linear learning is that in log-linear learning players have a fixed exponential clock rate of 1.

where we used that $Y \leq (N+1)^{A-1}$. We deduce that for $\beta = \Omega(\frac{1}{\max\{\epsilon,\Delta\}} \log(\frac{N^A}{\epsilon}))$, it holds that:

$$\mathbb{E}_{x\sim\mu_m}[\Phi(x)] \geq (1-\epsilon) \max_{x\in\Psi^{\mathcal{A}}} \Phi(x) \geq \max_{x\in\Psi^{\mathcal{A}}} \Phi(x) - \epsilon.$$

The proof now follows from the same analysis as in the proof of Theorem 3 in (Shah and Shin, 2010) with the exception that we do not make use of Lemma 6 in (Shah and Shin, 2010) but replace it with our analysis above. Concretely, we set β as specified in Equation (22) rather than as in (**?**)Eq. (8)]shah2010dynamics. ☐

C. Binary log-linear learning

Binary log-linear learning induces an irreducible and aperiodic Markov chain $\{X_t\}_{t\in\mathbb{Z}_+}$ with a time-reversible transition matrix $P \in \mathbb{R}^{\mathcal{A}\times\mathcal{A}}$ given by:

$$P_{a,\tilde{a}} = \frac{1}{N}\frac{1}{A}\frac{e^{\beta U_i(\tilde{a}_i,\tilde{a}_{-i})}}{e^{\beta U_i(a_i,\tilde{a}_{-i})} + e^{\beta U_i(\tilde{a}_i,\tilde{a}_{-i})}} \mathbb{1}_{\tilde{a}\in\mathcal{N}(a)} \tag{23}$$

where $\mathcal{N}(a) = \{\tilde{a} \in \mathcal{A}^N \mid \exists i \in [N] : \tilde{a}_{-i} = a_{-i}\}$. The additional term $1/A$ stems from the fact that player i first randomly samples an action $\tilde{a}_i$ and then decides between this action and her previous action. Arslan et al. (2007) show that its stationary distribution $\mu \in \Delta(\mathcal{A}^N)$ is given by :

$$\mu(a) = \frac{e^{\beta\Phi(a)}}{\sum_{\tilde{a}\in A^N} e^{\beta\Phi(\tilde{a})}} \quad \forall a \in \mathcal{A}^N. \tag{24}$$

The stationary distribution of binary log-linear learning is the same as that of log-linear learning (Equation (5)). Thus, log-linear- and binary log-linear learning converge to an approximately efficient Nash equilibrium in the long run. We briefly outline the proof of Theorem 4.1 and then provide a detailed proof.

Proof outline. The proof follows from the same line of arguments as the proof of Theorem 3.1. In particular, the first step in the proof of Theorem 3.1 remains the same since binary log-linear learning has the same stationary distribution as log-linear learning. In the second step in the proof of Theorem 3.1, the main difference is that the transition matrix in (23) of binary log-linear learning differs from the transition matrix in (4) of log-linear learning. Thus, the log-Sobolev constant of binary log-linear can be lower-bounded as follows:

$$\rho(PP^*) \geq \frac{16\pi^2 A^N \mu_{\min} p_{\min}^3}{25N^2A^2} \geq \frac{2\pi^2 e^{-4\beta}}{25N^2A^5}, \tag{25}$$

while the log-Sobolev constant of log-linear can be lower-bounded as follows:

$$\rho(PP^*) \geq \frac{16\pi^2 A^N \mu_{\min} p_{\min}^3}{25N^2A^2} \geq \frac{16\pi^2 e^{-4\beta}}{25N^2A^5}.$$

Then, we use Lemma 2.2 to show that

$$\|\mu^t - \mu\|_{TV} \leq \epsilon/4$$

for $t \geq \frac{1}{\rho(PP^*)}(\log\log\frac{1}{\mu_{\min}} + 2\log\frac{4}{\epsilon})$ with $\rho(PP^*)$ lower-bounded as in Equation (25). ☐

Proof of Theorem 4.1. By Lemma 3.3, the log-Sobolev constant $\rho(PP^*)$ can be lower-bounded as:

$$\rho(PP^*) \geq \frac{16\pi^2 A^N \mu_{\min} p_{\min}^3}{25N^2A^2} \geq \frac{2\pi^2 e^{-4\beta}}{25N^2A^5},$$

where we used that by definition of P in (23) and μ in (24) $\mu_{\min}$ and $p_{\min}$ can be lower-bounded as follows:

$$\mu_{\min} = \min_{a\in\mathcal{A}^N} \mu(a) \geq \frac{e^{-\beta}}{A^N}$$

$$P_{a,\tilde{a}} \geq \frac{e^{-\beta}}{N2A}, \quad \forall \tilde{a} \in \mathcal{A}^N(a) \;\Rightarrow\; p_{\min} = \frac{e^{-\beta}}{2A}.$$

Equation (2) in Lemma 2.2 provides the following upper bound on the mixing time:

$$t_{\mathrm{mix}}^{P}(\epsilon/4) \leq \frac{1}{\rho(PP^{*})}\left(\log\log\frac{1}{\mu_{*}} + 2\log\frac{4}{\epsilon}\right).$$

Plugging the bound on the log-Sobolev constant into this equation we obtain:

$$\begin{aligned} t_{\mathrm{mix}}^{P}(\epsilon/4) &\leq \frac{25N^2A^5}{2\pi^2}e^{4\beta}\left(\log\log\frac{1}{\mu_{\min}} + 2\log\frac{4}{\epsilon}\right) \\ &\leq \frac{25N^2A^5}{2\pi^2}e^{4\beta}\left(\log\log\frac{A^N}{e^{-\beta}} + 2\log\frac{4}{\epsilon}\right) \\ &\leq \frac{25N^2A^5}{2\pi^2}e^{4\beta}\left(\log\log A^N + \log\beta + 2\log\frac{4}{\epsilon}\right). \end{aligned}$$

Set t as:

$$t \geq \frac{25N^2A^5}{2\pi^2}e^{4\beta}\left(\log\log A^N + \log\beta + 2\log\frac{4}{\epsilon}\right) \tag{26}$$

and set β as:

$$\beta \geq \frac{1}{\max\{\epsilon/2, \Delta\}}\log\left((A^N - A^N(\epsilon/2))\left(\frac{4}{\epsilon A^N(\epsilon/2)} - \frac{1}{A^N(\epsilon/2)}\right)\right). \tag{27}$$

Then, we obtain the following upper bound:

$$\begin{aligned} \mathbb{E}[\Phi(a^t)] &= \mathbb{E}_{a\sim\mu^t}[\Phi(a)] \\ &\geq \mathbb{E}_{a\sim\mu}[\Phi(a)] - 2\|\mu^t - \mu\|_{TV}\max_{a\in\mathcal{A}^N}\Phi(a) \\ &\geq \max_{a\in\mathcal{A}^N}\Phi(a) - \frac{\epsilon}{2} - \frac{2\epsilon}{4} \\ &= \max_{a\in\mathcal{A}^N}\Phi(a) - \epsilon, \end{aligned}$$

where the third line follows from Lemma 3.2, the fact that $\|\mu^t - \mu\|_{TV} \leq \epsilon/4$ for t set as in Equation (26), and the fact that $\Phi(\cdot) \in [0, 1]$. Lemma 3.2 is applicable when all players adhere to binary-based log-linear learning rather than log-linear learning since the proof of Lemma 3.2 depends only on the stationary distribution μ of the corresponding learning rule which is the same for log-linear learning and binary log-linear learning. This concludes the proof of Theorem 4.1. □

D. Robustness of log-linear learning

In this section, we prove Theorem 4.3 and apply it to the corrupted-utility case to prove Corollary 4.4.

D.1. Proof of Theorem 4.3

Proof of Theorem 4.3. Consider a learning rule P, to prove Theorem 4.3, we first provide a decomposition that relates the expected value of the potential when the agents follow P defined in Equation (8) to the same quantity when the agents follow P_ℓ defined in Equation (4) instead.

We have for all $t, t' \in \mathbb{N}$ that:

$$\begin{aligned} \mathbb{E}_{\mu_0 P^t}[\Phi] &= \mathbb{E}_{\mu_0 P_\ell^{t'}}[\Phi] + \mathbb{E}_{\mu_0 P^t}[\Phi] - \mathbb{E}_{\mu_0 P_\ell^{t'}}[\Phi] \\ &\geq \mathbb{E}_{\mu_0 P_\ell^{t'}}[\Phi] - \sqrt{A^N}\|P^t - P_\ell^{t'}\|_2 \end{aligned} \tag{28}$$

where the last line follows because $|\Phi(a)| \leq 1$ for all $a \in \mathcal{A}^N$ and because $\|.\|_1 \leq \sqrt{A^N}\|\cdot\|_2$.

The rest of the proof is based on controlling $\|P^t - P_\ell^{t'}\|_2$ using our mixing time results.

Decomposition: Using Lemma 4.2 we obtain:

$$\begin{aligned}\|P^t - P_\ell^{t'}\|_2 &\leq \|P^t - \mu\|_2 + \|P_\ell^{t'} - \mu_\ell\|_2 + \|\mu - \mu_\ell\|_2 \\ &\leq \|P^t - \mu\|_2 + \|P_\ell^{t'} - \mu_\ell\|_2 + L(P_\ell)\|P - P_\ell\|_2 \\ &\leq 2\|P^t - \mu\|_{TV} + \|P_\ell^{t'} - \mu_\ell\|_2 + L(P_\ell)\|P - P_\ell\|_2,\end{aligned}$$

where $L(P_\ell) = \frac{2A^N}{\rho(P_\ell P_\ell^*)}(\log\log\frac{1}{\mu_{\ell,\min}} + \log(8A^N))$ follows from Lemma 4.2. In Theorem 3.1, we showed that $\mu_{\ell,\min} \geq \frac{e^{-\beta}}{A^N}$ and $\rho(P_\ell P_\ell^*) \geq \frac{16\pi^2 e^{-4\beta}}{25N^2A^5}$, therefore

$$L(P_\ell) \leq \frac{25N^2A^{N+5}e^{4\beta}}{8\pi^2}(\log\log A^N e^\beta + \log(8A^N)).$$

Thus, for

$$\begin{cases} t & \geq t_{\text{mix}}^P(\epsilon/(4\sqrt{A^N}))) \\ t' & \to \infty \\ \beta & = \log\left((A^N - A^N(\epsilon/2))\left(\frac{4}{\epsilon/2A^N(\epsilon/2)} - \frac{1}{A^N(\epsilon/2)}\right)\right)/\max\{\epsilon/2, \Delta\} \end{cases}$$

we have

$$\begin{cases} \|P^t - P_\ell^t\|_2 & \leq \epsilon/\left(2\sqrt{A^N}\right) + L(P_\ell)\|P - P_\ell\|_2 \\ \mathbb{E}_{a\sim\mu^0 P_\ell^{t'}}[\Phi(a)] & \geq \max_{a\in\mathcal{A}^N}\Phi(a) - \epsilon/2, \\ L(P_\ell) & = \mathcal{O}\left(N^2A^{N+5}e^{\frac{\log(A^N/\epsilon)}{\max\{\epsilon,\Delta\}}}\left(\log\log A^N e^{\frac{\log(A^N/\epsilon)}{\max\{\epsilon,\Delta\}}} + \log(A^N)\right)\right), \end{cases}$$

where the second line follows from Lemma 3.2. Plugging the above inequalities with the bound on L from Lemma 4.2 into the decomposition (28) proves the desired result.

We now bound the mixing time $t_{\text{mix}}^P(\epsilon/(4\sqrt{A^N}))$ of the Markov chain induced by P.

Mixing time bound: To bound the mixing time of the Markov chain induced by P, we use inequality (2) and Lemma 3.3. Assuming a lower bound of $p_{\min}/N$ on the probabilities of all feasible transitions implies a bound on the stationary distribution as we will show next.

Lower bound $(\mu_P)_{\min}$: Since P has a positive probability of transitioning from $a \in \mathcal{A}^N$ to any $\tilde{a} \in \mathcal{N}(a)$, it follows that the corresponding N-step transition P^N has a positive probability of transitioning from any $a \in \mathcal{A}^N$ to any $a' \in \mathcal{A}^N$, *i.e.,*

$$\forall a, a' \in \mathcal{A}^N: \quad P_{a,a'}^N \geq N!\,(p_{\min}/N)^N.$$

Note that the least probable transitions are when a and a' are such that $\forall i \in [N]: \ a_i \neq a_i'$. For all such transitions, the possible paths using P^N are the permutations of $\{1, \ldots, N\}$ (each of the N steps is a new player updating their action). There are $N!$ such permutations and each player $i \in [N]$ can update a_i to a_i' with probability larger than $p_{\min}/N$.

Since P is an irreducible and aperiodic transition matrix, the Markov chain induced by P has a unique stationary distribution μ_P. It is known that the Markov chain induced by P^N has the same stationary distribution as the one induced by P. Therefore, we have for all $a \in \mathcal{A}^N$:

$$\begin{aligned}\mu_P(a) &= \sum_{\tilde{a}\in\mathcal{A}^N} P_{\tilde{a},a}^N \mu_P(\tilde{a}) \\ &\geq \sum_{\tilde{a}\in\mathcal{A}^N} N!\,(p_{\min}/N)^N \mu_P(\tilde{a}) = N!\,(p_{\min}/N)^N.\end{aligned}$$

Therefore, $(\mu_P)_{\min} \geq N!\,(p_{\min}/N)^N$.

Deducing the mixing-time bound: We can now give an explicit bound on the mixing time of the transition P. First, we have by Lemma 3.3:

$$\begin{aligned}\rho(PP^*) &\geq \frac{16\pi^2 A^N (\mu_P)_{\min} p_{\min}^3}{25N^2} \\ &\geq \frac{4\pi^2 A^N p_{\min}^{N+3} N!}{25N^{N+2}}.\end{aligned}$$

Then, using Stirling's formula, we have $N! \geq \sqrt{2\pi N}\left(\frac{N}{e}\right)^N$, therefore

$$\rho(PP^*) \geq \frac{(2\pi)^{5/2} A^N p_{\min}^{N+3}}{25N^{3/2}e^N}$$

Then, using inequality (2), we have:

$$\begin{aligned}t_{\text{mix}}^P(\epsilon/(4\sqrt{A^N})) &\leq \frac{1}{\rho(PP^*)}\left(\log\log\frac{1}{(\mu_P)_{\min}} + 2\log\frac{4\sqrt{A^N}}{\epsilon}\right) \\ &\leq \frac{25N^{3/2}e^N}{(2\pi)^{5/2}A^N p_{\min}^{N+3}}\left(\log\log\frac{e^N}{p_{\min}^N\sqrt{2\pi N}} + 2\log\frac{4\sqrt{A^N}}{\epsilon}\right).\end{aligned}$$

This concludes the proof of Theorem 4.3. □

D.2. Proof of Corollary 4.4

The key idea is to show that the transition matrix of the Markov chain induced by corrupted utilities is close to its corruption-free counterpart.

Proof of Corollary 4.4. If all players adhere to log-linear learning with corrupted utilities, the induced Markov chain's transition matrix $\hat{P}$ is given, for all $a, \tilde{a} \in \mathcal{A}^N$ by:

$$\begin{aligned}\hat{P}_{a,\tilde{a}} &= \frac{1}{N}\frac{e^{\beta\hat{U}_i(\tilde{a}_i,\tilde{a}_{-i})}}{\sum_{a_i'\in\mathcal{A}_i} e^{\beta\hat{U}_i(a_i',\tilde{a}_{-i})}}\mathbb{1}_{\tilde{a}\in\mathcal{N}(a)}, \\ &= \frac{1}{N}\frac{e^{\beta(U_i(\tilde{a}_i,\tilde{a}_{-i})+\xi_i(\tilde{a}_i,\tilde{a}_{-i}))}}{\sum_{a_i'\in\mathcal{A}_i} e^{\beta(U_i(a_i',\tilde{a}_{-i})+\xi_i(a_i',\tilde{a}_{-i}))}}\mathbb{1}_{\tilde{a}\in\mathcal{N}(a)}.\end{aligned}$$

Since we assumed that the noise is bounded, we can deduce that

$$P_{a,\tilde{a}}e^{-2\beta\xi} \leq P_{a,\tilde{a}} \leq P_{a,\tilde{a}}e^{2\beta\xi},$$

where $P_{a,\tilde{a}} = \frac{1}{N}\frac{e^{\beta U_i(\tilde{a}_i,\tilde{a}_{-i})}}{\sum_{a_i'\in\mathcal{A}_i} e^{\beta U_i(a_i',\tilde{a}_{-i})}}\mathbb{1}_{\tilde{a}\in\mathcal{N}(a)}$ is the transition with the noise-free utility. This entails that

$$P_{a,\tilde{a}}(e^{-2\beta\xi} - 1) \leq \hat{P}_{a,\tilde{a}} - P_{a,\tilde{a}} \leq P_{a,\tilde{a}}(e^{2\beta\xi} - 1),$$

then, since $e^{-2\beta\xi} - 1 < 0$ and $P_{a,\tilde{a}} \leq 1/N$ for all $a, \tilde{a} \in \mathcal{A}^N$, we deduce that

$$(e^{-2\beta\xi} - 1)/N \leq \hat{P}_{a,\tilde{a}} - P_{a,\tilde{a}} \leq (e^{2\beta\xi} - 1)/N,$$

and

$$|\hat{P}_{a,\tilde{a}} - P_{a,\tilde{a}}| \leq \frac{1}{N}\max\left\{e^{2\beta\xi} - 1, 1 - e^{-2\beta\xi}\right\},$$

Finally, since $2\beta\xi \leq 1$ and by using that: $1 - e^{-x} < x$ for $x > 0$, and that: $e^x - 1 < \frac{7}{4}x$ for $x \in [0, 1]$. Then,

$$|\hat{P}_{a,\tilde{a}} - P_{a,\tilde{a}}| \leq \frac{1}{N}\max\left\{\frac{7}{2}\beta\xi, 2\beta\xi\right\} = \frac{7}{2N}\beta\xi,$$

and finally

$$\|\hat{P} - P\|_2 \leq \sqrt{\sum_{a,\tilde{a} \in \mathcal{A}^N} \frac{49}{4N^2}\beta^2\xi^2} = \frac{7A^N}{2N}\beta\xi.$$

Also, since $P_{a,\tilde{a}} \geq P_{a,\tilde{a}} e^{-2\beta\xi}$ and using $P_{a,\tilde{a}} \geq \frac{e^{-\beta}}{NA}$ then we deduce that $P_{a,\tilde{a}} \geq \frac{e^{-\beta(1+2\xi)}}{NA}$. We conclude the proof with a straightforward application of Theorem 4.3 with $p_{\min} = e^{-\beta(1+2\xi)}/A$ and $\|\hat{P} - P\|_2 \leq \frac{7A^N}{2N}\beta\xi$. □

D.3. Proof of Corollary 4.5

Similar to Corollary 4.4, we proceed by showing that the transition matrix of the Markov chain induced by fixed-share log-linear learning is close to that of log-linear learning.

Proof of Corollary 4.5. If all players adhere to fixed-share log-linear learning, the induced Markov chain's transition matrix $\hat{P}$ is given, for all $a, \tilde{a} \in \mathcal{A}^N$ by:

$$\hat{P}_{a,\tilde{a}} = \frac{1}{N}\left(\frac{\xi}{A} + \frac{(1-\xi)e^{\beta U_i(\tilde{a}_i,\tilde{a}_{-i})}}{\sum_{a_i' \in \mathcal{A}} e^{\beta U_i(a_i',\tilde{a}_{-i})}}\right)\mathbb{1}_{\tilde{a}\in\mathcal{N}(a)}. \tag{29}$$

Then, we have that

$$\hat{P}_{a,\tilde{a}} \geq \left(\frac{\xi}{NA} + \frac{(1-\xi)e^{-\beta}}{NA}\right)\mathbb{1}_{\tilde{a}\in\mathcal{N}(a)},$$

which entails that $\hat{P}$ satisfies the condition of Theorem 4.3 with $p_{\min} \geq \frac{\xi}{A} + \frac{(1-\xi)e^{-\beta}}{A}$.

Additionally, we can show that:

$$\begin{aligned}
\hat{P}_{a,\tilde{a}} - P_{a,\tilde{a}} &= \frac{1}{N}\left(\frac{\xi}{A} - \frac{\xi e^{\beta U_i(\tilde{a}_i,\tilde{a}_{-i})}}{\sum_{a_i' \in \mathcal{A}} e^{\beta U_i(a_i',\tilde{a}_{-i})}}\right)\mathbb{1}_{\tilde{a}\in\mathcal{N}(a)} \\
&= \frac{\xi}{N}\left(\frac{1}{A} - \frac{e^{\beta U_i(\tilde{a}_i,\tilde{a}_{-i})}}{\sum_{a_i' \in \mathcal{A}} e^{\beta U_i(a_i',\tilde{a}_{-i})}}\right)\mathbb{1}_{\tilde{a}\in\mathcal{N}(a)},
\end{aligned}$$

where P is the transition matrix of log-linear learning.

Therefore,

$$\begin{aligned}
\sum_{a,\tilde{a}\in\mathcal{A}^N}\left(\hat{P}_{a,\tilde{a}} - P_{a,\tilde{a}}\right)^2 &= \frac{\xi^2}{N^2}\sum_{a,\tilde{a}\in\mathcal{A}^N}\left(\frac{1}{A^2} - \frac{2e^{\beta U_i(\tilde{a}_i,\tilde{a}_{-i})}}{A\sum_{a_i'\in\mathcal{A}} e^{\beta U_i(a_i',\tilde{a}_{-i})}} + \frac{e^{2\beta U_i(\tilde{a}_i,\tilde{a}_{-i})}}{\left(\sum_{a_i'\in\mathcal{A}} e^{\beta U_i(a_i',\tilde{a}_{-i})}\right)^2}\right)\mathbb{1}_{\tilde{a}\in\mathcal{N}(a)} \\
&\leq \frac{\xi^2}{N^2}\sum_{a\in\mathcal{A}^N}\left(\frac{N}{A} - \frac{2N}{A} + N\right)\mathbb{1}_{\tilde{a}\in\mathcal{N}(a)} \\
&\leq NA^N,
\end{aligned}$$

where the second line follows because from any action profile $a \in A^N$, there are NA possible transitions (A possible actions times N possible player selections). We also used $\sum_{\tilde{a}\in\mathcal{A}^N} \frac{e^{\beta U_i(\tilde{a}_i,\tilde{a}_{-i})}}{\sum_{a_i'\in\mathcal{A}} e^{\beta U_i(a_i',\tilde{a}_{-i})}}\mathbb{1}_{\tilde{a}\in\mathcal{N}(a)} = 1$ and that$\sum_{\tilde{a}\in\mathcal{A}^N} \frac{e^{\beta U_i(\tilde{a}_i,\tilde{a}_{-i})}}{\left(\sum_{a_i'\in\mathcal{A}} e^{\beta U_i(a_i',\tilde{a}_{-i})}\right)^2}\mathbb{1}_{\tilde{a}\in\mathcal{N}(a)} \leq 1$.

Finally, since the spectral norm is smaller than the Frobenius norm, then

$$\begin{aligned}\|\hat{P}-P\|_2 &\leq \sqrt{\sum_{a,\tilde{a}\in\mathcal{A}^N}\left(\hat{P}_{a,\tilde{a}}-P_{a,\tilde{a}}\right)^2}\\ &\leq \xi\sqrt{\frac{A^N}{N}}.\end{aligned}$$

The proof is then concluded by a straightforward application of Theorem 4.3 with $p_{\min} \geq \frac{\xi}{A} + \frac{(1-\xi)e^{-\beta}}{A}$ and $\|\hat{P}-P\|_2 \leq \xi\sqrt{\frac{A^N}{N}}$.

□

D.4. Proof of Lemma 4.2

Proof of Lemma 4.2. Denote by $M \in \mathbb{R}^{A^N \times A^N}$ the matrix, where each row corresponds to μ_1. For all $t \in \mathbb{N}$, we have that:

$$\begin{aligned}\mu_1-\mu_2 &= (P_1^t)^\top(\mu_1-\mu_2) + (P_1^t-P_2^t)^\top\mu_2\\ &= (P_1^t-M)^\top(\mu_1-\mu_2) + M^\top(\mu_1-\mu_2)\\ &\quad + (P_1^t-P_2^t)^\top\mu_2.\end{aligned}$$

This yields:

$$\begin{aligned}\|\mu_1-\mu_2\|_2 &\leq \|(P_1^t-M)^\top(\mu_1-\mu_2)\|_2\\ &\quad + \|M^\top(\mu_1-\mu_2)\|_2 + \|(P_1^t-P_2^t)^\top\|_2\|\mu_2\|_2\\ &\leq \|P_1^t-M\|_2\|\mu_1-\mu_2\|_2\\ &\quad + \|M^\top(\mu_1-\mu_2)\|_2 + \|P_1^t-P_2^t\|_2\\ &\leq 2\sqrt{A^N}\|P_1^t-M\|_{TV}\|\mu_1-\mu_2\|_2\\ &\quad + \|M^\top(\mu_1-\mu_2)\|_2 + \|P_1^t-P_2^t\|_2\end{aligned}$$

where in the last inequality we used the equivalence of $\|\cdot\|_2$ and $\|\cdot\|_1$ and that $\|\cdot\|_1 = 2\|\cdot\|_{TV}$ by definition of the total variation distance. Furthermore:

$$\begin{aligned}M^\top(\mu_1-\mu_2) &= \Bigg(\mu_1(a)\underbrace{\sum_{a'\in\mathcal{A}^N}(\mu_1(a')-\mu_2(a'))}_{=0}\Bigg)_{a\in\mathcal{A}^N}\\ &= 0.\end{aligned}$$

Therefore, we obtain that:

$$\begin{aligned}\|\mu_1-\mu_2\|_2 &\leq 2\sqrt{A^N}\|P_1^t-M\|_{TV}\|\mu_1-\mu_2\|_2\\ &\quad + \|P_1^t-P_2^t\|_2. \end{aligned} \tag{30}$$

Note that for the second term on the right-hand side in the equation above we have:

$$\begin{aligned}P_1^t-P_2^t &= P_1^t + \sum_{l=1}^{t-1}(P_1^{t-l}P_2^l - P_1^{t-l}P_2^l) - P_2^t\\ &= \sum_{l=1}^{t}(P_1^{t-l+1}P_2^{l-1} - P_1^{t-l}P_2^l)\\ &= \sum_{l=1}^{t}(P_1^{t-l}(P_1-P_2)P_2^{l-1}).\end{aligned}$$

By applying the norm operator we find that:

$$\begin{aligned}\|P_1^t - P_2^t\|_2 &\leq \sum_{l=1}^{t} \|P_1^{t-l}\|_2 \|P_1 - P_2\|_2 \|P_2^{l-1}\|_2 \\ &\leq tA^N \|P_1 - P_2\|_2,\end{aligned}$$

since $\|P\|_2 \leq \sqrt{A^N}$ holds for all transition matrices P over $\mathcal{A}^N$, and in particular for P_1^{t-l} and P_2^{l-1}. By plugging the above into inequality (30) we find:

$$\begin{aligned}\|\mu_1 - \mu_2\|_2 \leq\ & 2\sqrt{A^N} \|P_1^t - M\|_{TV} \|\mu_1 - \mu_2\|_2 \\ &+ tA^N \|P_1 - P_2\|_2.\end{aligned}$$

Finally, by choosing $t = t_{\text{mix}}\left(1/\sqrt{16A^N}\right)$ we find:

$$\|\mu_1 - \mu_2\|_2 \leq 2t_{\text{mix}}\left(1/\sqrt{16A^N}\right) A^N \|P_1 - P_2\|_2.$$

The proof is then concluded by using the mixing-time bound from inequality (2). □