SUPPLEMENTARY MATERIALS: FINITE-SAMPLE GUARANTEES FOR LEARNING DYNAMICS IN ZERO-SUM POLYMATRIX GAMES*

FATHIMA ZARIN FAIZAL[†], ASUMAN OZDAGLAR[†], AND MARTIN J. WAINWRIGHT[†‡]

SM1. The full information case. In this section, we first state our results for the inverse polynomial stepsize schedule in Subsection SM1.1. We then complete the proof of Theorem 3.1 for this case along with the inverse linear case.

SM1.1. The inverse polynomial stepsize. In the following lemma, we state an upper bound on the Nash gap and a bound on the iteration complexity for an inverse polynomial stepsize schedule.

LEMMA SM1.1. *Consider the full information dynamics (Algorithm 2.1) initialized with $\boldsymbol{\pi}_1$ and $\tau > 0$. For the inverse polynomial stepsize sequence $\beta_k = \frac{\beta}{(k+k_0)^\eta}$ with $\beta, \eta \in (0,1)$ and $k_0 \geq \lceil \left(\frac{1-\eta}{\beta}\right)^{1/\eta} \rceil$, we have the following upper bound on the Nash gap at any iteration K:*

$$\mathrm{NG}(\boldsymbol{\pi}_{K+1}) \leq \exp\big(-\frac{\beta}{1-\eta}\big((K+k_0+1)^{1-\eta} - (1+k_0)^{1-\eta}\big)\big)V_1 + N\tau \log A_{\max} + \frac{2N\|\boldsymbol{R}\|_2^2 \beta}{\tau(K+k_0)^\eta}.$$

When the temperature parameter $\tau = c_1\epsilon/(N\log A_{\max})$, the iteration complexity $K(\epsilon)$ is bounded as:

$$K(\epsilon) \leq \big(k_0^{1-\eta} + \frac{(1-\eta)}{\beta}\log\frac{V_1}{\epsilon}\big)^{1/(1-\eta)} + \big(\frac{c_2 N^2 \|\boldsymbol{R}\|_2^2 \beta \log A_{\max}}{\epsilon^2}\big)^{1/\eta}.$$

SM1.2. Auxiliary calculations for Theorem 3.1 and Corollary 3.2. In the main text, we used the drift inequality (4.1) to prove the bound in Theorem 3.1 for constant stepsizes. Here we cover the two remaining cases.

SM1.2.1. Inverse linear stepsizes. Consider the inverse linear stepsizes $\beta_k = \frac{\beta}{k}$ for some $\beta > 0$. Solving the inequality (4.1) gives that $V_k := \mathcal{V}(\boldsymbol{\pi}_k)$ satisfies the equation

$$V_{K+1} \leq \prod_{k=1}^{K}(1-\beta_k)V_1 + \frac{N\|\boldsymbol{R}\|_2^2}{\tau}\sum_{k=1}^{K}\beta_k^2 \prod_{j=k+1}^{K}(1-\beta_j) + N\tau\log A_{\max}\sum_{k=1}^{K}\beta_k\prod_{j=k+1}^{K}(1-\beta_j). \tag{SM1.1}$$

*Submitted to the editors DATE.

[†]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA (fathima@mit.edu,asuman@mit.edu).

[‡]Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA (mjwain@mit.edu).

Bounding the first term:

$$\prod_{k=1}^{K}(1-\beta_k) \;=\; \exp\big(\sum_{k=1}^{K}\log(1-\beta_k)\big) \overset{(a)}{\leq} \exp\big(-\sum_{k=1}^{K}\beta_k\big)$$
$$\overset{(b)}{\leq} \exp\big(-\beta\int_1^{K+1}\frac{1}{x}\,dx\big)$$
$$= \big(\frac{1}{K+1}\big)^{\beta},$$

where in step (a), we used the bound $\log(1-x) \leq -x$ for $x \in (0,1)$ and in step (b), we bounded the Riemann sum.

Bounding the second term: Using a series of arguments similar to the upper bound for the first term yields

$$\prod_{j=k+1}^{K}(1-\beta_j) = \exp\big(\sum_{j=k+1}^{K}\log(1-\beta_j)\big)$$
$$\leq \exp\big(-\sum_{j=k+1}^{K}\beta_j\big)$$
$$\leq \exp\big(-\beta\int_{k+1}^{K+1}\frac{1}{x}\,dx\big)$$
$$= \big(\frac{k+1}{K+1}\big)^{\beta}. \tag{SM1.2}$$

Substituting this bound into the expression for the second term yields

$$\frac{N\|\boldsymbol{R}\|_2^2}{\tau}\sum_{k=1}^{K}\beta_k^2\prod_{j=k+1}^{K}(1-\beta_j) \leq \frac{N\|\boldsymbol{R}\|_2^2}{\tau}\sum_{k=1}^{K}\beta_k^2\big(\frac{k+1}{K+1}\big)^{\beta}$$
$$\leq \frac{4N\|\boldsymbol{R}\|_2^2\beta^2}{\tau(K+1)^{\beta}}\sum_{k=1}^{K}\frac{1}{(k+1)^{2-\beta}}.$$

We now consider five cases.

Case 1. If $\beta = 1$, then we have

$$\sum_{k=1}^{K}\frac{1}{(k+1)^{2-\beta}} \leq \int_1^{K+1}\frac{1}{x}\,dx = \log(K+1).$$

Therefore,

$$\frac{4N\|\boldsymbol{R}\|_2^2\beta^2}{\tau(K+1)^{\beta}}\sum_{k=1}^{K}\frac{1}{(k+1)^{2-\beta}} \leq \frac{4N\|\boldsymbol{R}\|_2^2\log(K+1)}{\tau(K+1)}.$$

Case 2. If $\beta > 2$, then we can write

$$\sum_{k=1}^{K}\frac{1}{(k+1)^{2-\beta}} = \sum_{k=2}^{K+1}k^{\beta-2} \leq \int_2^{K+2}x^{\beta-2}dx \leq \frac{(K+2)^{\beta-1}}{\beta-1}.$$

Therefore,

$$\begin{aligned}\frac{4N\|\boldsymbol{R}\|_2^2\beta^2}{\tau(K+1)^\beta}\sum_{k=1}^{K}\frac{1}{(k+1)^{2-\beta}} &\le \frac{4N\|\boldsymbol{R}\|_2^2\beta^2(K+2)^{\beta-1}}{\tau(\beta-1)(K+1)^\beta}\\ &\le \frac{4N\|\boldsymbol{R}\|_2^2\beta^2 2^{\beta-1}}{\tau(\beta-1)(K+1)}.\end{aligned}$$

Case 3. If $\beta = 2$, then

$$\sum_{k=1}^{K}\frac{1}{(k+1)^{2-\beta}} = K.$$

Therefore,

$$\begin{aligned}\frac{4N\|\boldsymbol{R}\|_2^2\beta^2}{\tau(K+1)^\beta}\sum_{k=1}^{K}\frac{1}{(k+1)^{2-\beta}} &\le \frac{4N\|\boldsymbol{R}\|_2^2\beta^2 K}{\tau(K+1)^\beta}\\ &\le \frac{16N\|\boldsymbol{R}\|_2^2}{\tau(K+1)}.\end{aligned}$$

Case 4. If $\beta \in (1,2)$, then

$$\sum_{k=1}^{K}\frac{1}{(k+1)^{2-\beta}} = \sum_{k=2}^{K+1}\frac{1}{k^{2-\beta}} \le \int_1^{K+1}\frac{1}{x^{2-\beta}}dx \le \frac{K^{\beta-1}}{\beta-1}.$$

Therefore, we have

$$\begin{aligned}\frac{4N\|\boldsymbol{R}\|_2^2\beta^2}{\tau(K+1)^\beta}\sum_{k=1}^{K}\frac{1}{(k+1)^{2-\beta}} &\le \frac{4N\|\boldsymbol{R}\|_2^2\beta^2 K^{\beta-1}}{\tau(\beta-1)(K+1)^\beta}\\ &\le \frac{4N\|\boldsymbol{R}\|_2^2\beta^2}{\tau(\beta-1)K}.\end{aligned}$$

Case 5. Finally, if $\beta \in (0,1)$, then we have

$$\sum_{k=1}^{K}\frac{1}{(k+1)^{2-\beta}} = \sum_{k=2}^{K+1}\frac{1}{k^{2-\beta}} \le \int_1^{K+1}\frac{1}{x^{2-\beta}}dx \le \frac{1}{1-\beta}.$$

Therefore,

$$\frac{4N\|\boldsymbol{R}\|_2^2\beta^2}{\tau(K+1)^\beta}\sum_{k=1}^{K}\frac{1}{(k+1)^{2-\beta}} \le \frac{4N\|\boldsymbol{R}\|_2^2\beta^2}{\tau(1-\beta)(K+1)^\beta}.$$

<u>Bounding the third term:</u> Using equation (SM1.2) and a series of arguments similar to the upper bound for the second term:

$$\begin{aligned}N\tau\log A_{\max}\sum_{k=1}^{K}\beta_k\prod_{j=k+1}^{K}(1-\beta_j) &\le N\tau\log A_{\max}\sum_{k=1}^{K}\beta_k\Big(\frac{k+1}{K+1}\Big)^\beta\\ &\le \frac{2N\tau\log A_{\max}\beta}{(K+1)^\beta}\sum_{k=1}^{K}\frac{1}{(k+1)^{1-\beta}}.\end{aligned}$$

Case 1. If $\beta = 1$, then we have

$$\sum_{k=1}^{K} \frac{1}{(k+1)^{1-\beta}} = K.$$

Therefore,

$$\frac{2N\tau \log A_{\max}\beta}{(K+1)^{\beta}} \sum_{k=1}^{K} \frac{1}{(k+1)^{1-\beta}} \leq 2N\tau \log A_{\max}.$$

Case 2. If $\beta > 1$, then we can write

$$\sum_{k=1}^{K} \frac{1}{(k+1)^{1-\beta}} = \sum_{k=2}^{K+1} k^{\beta-1} \leq \int_{2}^{K+2} x^{\beta-1} dx \leq \frac{(K+2)^{\beta}}{\beta}.$$

Therefore,

$$\frac{2N\tau \log A_{\max}\beta}{(K+1)^{\beta}} \sum_{k=1}^{K} \frac{1}{(k+1)^{1-\beta}} = \frac{4\tau \log A_{\max}(K+2)^{\beta}}{(K+1)^{\beta}} \leq 2N\tau 2^{\beta} \log A_{\max}.$$

Case 3. If $\beta < 1$, then

$$\sum_{k=1}^{K} \frac{1}{(k+1)^{1-\beta}} = \sum_{k=2}^{K+1} \frac{1}{k^{1-\beta}} \leq \int_{1}^{K+1} \frac{1}{x^{1-\beta}} dx \leq \frac{(K+1)^{\beta}}{\beta}.$$

Therefore,

$$\frac{2N\tau \log A_{\max}\beta}{(K+1)^{\beta}} \sum_{k=1}^{K} \frac{1}{(k+1)^{1-\beta}} \leq 2N\tau \log A_{\max}.$$

The statement of Theorem 3.1 for inverse linear stepsizes is for the specific case when $\beta \in (1, 2]$. To bound the iteration complexity, we choose $\tau = \epsilon/(24N \log A_{\max})$.

SM1.2.2. Inverse polynomial stepsize. Now suppose that $\beta_K = \frac{\beta}{(K+k_0)^{\eta}}, \eta \in (0, 1)$ and $\beta \in (0, 1)$. Recall that V_{K+1} satisfies the inequality

$$\begin{aligned} V_{K+1} \leq \prod_{k=1}^{K} \big(1-\beta_k\big) V_1 + \frac{N\|\boldsymbol{R}\|_2^2}{\tau} \sum_{k=1}^{K} \beta_k^2 \prod_{j=k+1}^{K} (1-\beta_j) \\ + N\tau \log A_{\max} \sum_{k=1}^{K} \beta_k \prod_{j=k+1}^{K} (1-\beta_j). \end{aligned}$$

<u>Bounding the first term:</u> We have

$$\begin{aligned} \prod_{k=1}^{K} (1-\beta_k) &\leq \exp\big(-\beta \int_{1}^{K+1} \frac{1}{(x+k_0)^{\eta}}\, dx\big) \\ &= \exp\big(-\frac{\beta}{1-\eta}\big((K+k_0+1)^{1-\eta} - (1+k_0)^{1-\eta}\big)\big). \end{aligned}$$

Bounding the second term: We define the sequence $\{u_k\}_{k\geq 1}$ via the recursion

$$u_{k+1} = \big(1-\beta_k\big)u_k + \beta_k^2 \qquad \text{with initialization } u_1 = 0.$$

Unwrapping this recursion, we find that $u_K = \sum_{k=1}^{K} \beta_k^2 \prod_{j=k+1}^{K}(1-\beta_j)$. It can be shown by induction (see p.36, [SM1]) that $u_K \leq 2\beta_K$. Therefore,

$$\frac{N\|\boldsymbol{R}\|_2^2}{\tau}\sum_{k=1}^{K}\beta_k^2\prod_{j=k+1}^{K}(1-\beta_j) = \frac{N\|\boldsymbol{R}\|_2^2 u_K}{\tau} \leq \frac{2N\|\boldsymbol{R}\|_2^2\beta_K}{\tau} = \frac{2N\|\boldsymbol{R}\|_2^2\beta}{\tau(K+k_0)^\eta}.$$

Bounding the third term: We define the sequence $\{u_k\}_{k\geq 1}$ via the recursion

$$u_{k+1} = \big(1-\beta_k\big)u_k + \beta_k \quad \text{with initial value } u_1 = 0.$$

Note that since $\beta_k \in (0,1)$, it follows that $u_k \leq 1$ for all $k = 1, 2, \ldots$. Expanding out the recursion yields $u_K = \sum_{k=1}^{K} \beta_k \prod_{j=k+1}^{K}(1-\beta_j)$. Combining with the inequality $u_K \leq 1$, we find that

$$V_{K+1} \leq \exp\Big(-\frac{\beta}{1-\eta}\big((K+k_0+1)^{1-\eta} - (1+k_0)^{1-\eta}\big)\Big)V_1 + \frac{2N\|\boldsymbol{R}\|_2^2\beta}{\tau(K+k_0)^\eta} + N\tau\log A_{\max}.$$

The bound on the iteration complexity in Lemma SM1.1 then follows by setting $\tau = \epsilon/(3N\log A_{\max})$.

SM2. Proofs of key technical lemmas. We now collect together the proofs of the key technical lemmas that were used in the proof of Theorem 3.3 from Subsection 4.2. More specifically, we prove Lemmas 4.1, 4.2, 4.3 and 4.4 in Sections SM2.1, SM2.2, SM2.4, and SM2.5, respectively.

SM2.1. Proof of Lemma 4.1. We split our proof into parts, corresponding to the two statements in the lemma. In both cases, we make use of Danskin's theorem [SM3] to compute the gradient

$$\nabla_{\pi^i}\mathcal{V}(\boldsymbol{\pi}) = \sum_{j\in\mathcal{N}} R^{(j,i)\intercal}\sigma_\tau(q^j(\boldsymbol{\pi})). \tag{SM2.1}$$

Proof of part (a). In order to compute the Hessian, we take the derivative of the gradient from equation (SM2.1). Doing so via chain rule yields

$$\nabla^2_{\pi^i,\pi^m}\mathcal{V}(\boldsymbol{\pi}) = \frac{1}{\tau}\sum_{j\in\mathcal{N}} R^{(j,i)\intercal}\Sigma_\tau^j(\boldsymbol{\pi})R^{(j,m)},$$

where $\Sigma^j(\boldsymbol{\pi}) := \operatorname{diag}\big(\sigma_\tau(q^j(\boldsymbol{\pi})\big) - \sigma_\tau\big(q^j(\boldsymbol{\pi})\big)\sigma_\tau\big(q^j(\boldsymbol{\pi})\big)^\intercal$ for each $j \in \mathcal{N}$. The Hessian of $\mathcal{V}$ can now be decomposed in the following way:

$$\nabla^2\mathcal{V}(\boldsymbol{\pi}) = \frac{1}{\tau}\boldsymbol{R}^\intercal\boldsymbol{\Sigma}(\boldsymbol{\pi})\boldsymbol{R},$$

where $\boldsymbol{\Sigma}(\boldsymbol{\pi})$ is a block diagonal matrix with the i^{th} diagonal block $\Sigma^i(\boldsymbol{\pi})$ for $i \in \mathcal{N}$. By the sub-multiplicativity of the operator norm, we have

$$\|\nabla^2\mathcal{V}(\boldsymbol{\pi})\|_2 \leq \frac{1}{\tau}\|\boldsymbol{R}\|_2^2\ \|\boldsymbol{\Sigma}(\boldsymbol{\pi})\|_2.$$

Turning to the matrix $\Sigma^i(\boldsymbol{\pi})$, we recognize it as the covariance matrix of a multinomial random vector, so that it must be positive semi-definite (PSD). It follows that the matrix $\boldsymbol{\Sigma}(\boldsymbol{\pi})$ is also PSD, so that its operator norm can be written as $\|\boldsymbol{\Sigma}(\boldsymbol{\pi})\|_2 = \max_{j\in\mathcal{N}} \lambda_{\max}(\Sigma^j(\boldsymbol{\pi}))$. Now observe that

$$\Sigma^i(\boldsymbol{\pi}) \preceq \operatorname{diag}\big(\sigma_\tau(q^i(\pi^{-i}))\big),$$

where $\preceq$ denotes the PSD order. Thus, we have

$$\lambda_{\max}(\Sigma^i((\boldsymbol{\pi})) \leq \lambda_{\max}\Big[\operatorname{diag}\big(\Sigma^i(q^i(\pi^{-i}))\big)\Big] \leq 1.$$

Putting together the pieces yields the bound $\|\nabla^2\mathcal{V}(\boldsymbol{\pi})\|_2 \leq \frac{\|\boldsymbol{R}\|_2^2}{\tau}$, as claimed.

Proof of part (b). Again making use of the gradient representation (SM2.1) yields

$$\begin{aligned}\Big\langle \nabla_{\pi^i}\mathcal{V}(\boldsymbol{\pi}), \sigma_\tau(q^i(\pi^{-i})) - \pi^i\Big\rangle &= \Big\langle \sum_{j\in\mathcal{N}} R^{(j,i)\intercal}\sigma_\tau(q^j(\boldsymbol{\pi})), \sigma_\tau(q^i(\pi^{-i})) - \pi^i\Big\rangle \\ &= \sum_{j\in\mathcal{N}} \sigma_\tau(q^j(\boldsymbol{\pi}))^\intercal R^{(j,i)}\sigma_\tau(q^j(\boldsymbol{\pi})) - \sum_{j\in\mathcal{N}} \sigma_\tau(q^j(\boldsymbol{\pi}))^\intercal R^{(j,i)}\pi^i\end{aligned}$$

for each $i \in \mathcal{N}$. The zero-sum property (2.1) ensures that

$$\sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{N}} \sigma_\tau(q^j(\boldsymbol{\pi}))^\intercal R^{(j,i)}\sigma_\tau(q^i(\pi^{-i})) = 0.$$

Combined with the definition (2.9) of the Lyapunov function, we find that

$$\begin{aligned}\sum_{i\in\mathcal{N}}\Big\langle \nabla_{\pi^i}\mathcal{V}(\boldsymbol{\pi}), \sigma_\tau(q^i(\pi^{-i})) - \pi^i\Big\rangle &= -\sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{N}} \sigma_\tau(q^j(\boldsymbol{\pi}))^\intercal R^{(j,i)}\pi^i \\ &= -\sum_{j\in\mathcal{N}} \sigma_\tau(q^j(\boldsymbol{\pi}))^\intercal q^j(\boldsymbol{\pi}) \\ &= -\sum_{j\in\mathcal{N}} \sigma_\tau(q^j(\boldsymbol{\pi}))^\intercal q^j(\boldsymbol{\pi}) + \tau H(\sigma_\tau(q^j(\boldsymbol{\pi}))) \\ &\qquad - \tau H(\sigma_\tau(q^j(\boldsymbol{\pi}))) \\ &\leq -\mathcal{V}(\boldsymbol{\pi}) + N\tau\log A_{\max},\end{aligned}$$

which completes the proof.

SM2.2. Proof of Lemma 4.2. This section is devoted to the proof of the drift inequality stated in Lemma 4.2. It suffices to establish the following two claims:

(SM2.2a)

$$\mathbb{E}V_{k+1} \leq \big(1-\beta_k(1-r)\big)\mathbb{E}V_k + \tfrac{4N\|\boldsymbol{R}\|_2^2}{\tau}\beta_k^2 + \beta_k\tfrac{\|\boldsymbol{R}\|_2^2}{2r\tau^3}\mathbb{E}W_k + 2N\beta_k\tau\log A_{\max},$$

as well as

$$\text{(SM2.2b)}\quad \begin{aligned}\mathbb{E}W_{k+1} \leq \big((1-\alpha_k)^2 + \alpha_k^2\tfrac{A_{\max}N}{\delta}\big)\mathbb{E}W_k + 4N\beta_k^2\|\boldsymbol{R}\|_2^2 + \tfrac{4N^2A_{\max}\|\boldsymbol{R}\|_2^2}{\delta}\alpha_k^2 \\ + (1-\alpha_k)\tfrac{2\beta_k}{\tau r}\|\boldsymbol{R}\|_2^2\mathbb{E}W_k + (1-\alpha_k)\beta_k r\mathbb{E}V_k,\end{aligned}$$

for any choice of scalar $r \in (0, \frac{1}{2})$.

Given these claims, inequality (4.7) in Lemma 4.2 follows by summing together the inequalities (SM2.2a) and (SM2.2b). The rest of this section is devoted to proving each of these drift inequalities.

Proof of $\mathcal{V}$-inequality (SM2.2a). Our proof makes use of the previously stated result from Lemma 4.1 on the properties of the Lyapunov function $\mathcal{V}$. We also require some additional properties, which we summarize in the following auxiliary result:

LEMMA SM2.1. *The Lyapunov function $\mathcal{V}$ has the following properties:*

(a) We have the lower bound

$$\mathcal{V}(\boldsymbol{\pi}) \geq \frac{\tau}{2}\big\|\sigma_\tau(\boldsymbol{q}(\boldsymbol{\pi})) - \boldsymbol{\pi}\big\|_2^2 \qquad \text{for any mixed strategy } \boldsymbol{\pi} \in \boldsymbol{\Delta}.$$

(b) For any mixed strategy $\boldsymbol{\pi} \in \boldsymbol{\Delta}$ and $\boldsymbol{u} = (u^i)_{i\in\mathcal{N}} \in \otimes_{i\in\mathcal{N}}\mathbb{R}^{|\mathcal{A}^i|}$, we have

$$\big\langle \nabla_{\boldsymbol{\pi}}\mathcal{V}(\boldsymbol{\pi}), \sigma_\tau(\boldsymbol{u}) - \sigma_\tau(\boldsymbol{R}\boldsymbol{\pi})\big\rangle \leq N\tau\log A_{\max} + r\mathcal{V}(\boldsymbol{\pi}) + \frac{\|\boldsymbol{R}\|_2^2}{2r\tau^3}\|\boldsymbol{q}(\boldsymbol{\pi}) - \boldsymbol{u}\|_2^2,$$

valid for any constant $r \in (0,1)$.

See subsection SM3.1 for the proof of this auxiliary claim.

From the proof of Theorem 3.1, recall inequality (4.4), which ensures that

$$V_{k+1} \leq V_k + \beta_k\big\langle \nabla_{\boldsymbol{\pi}}\mathcal{V}(\boldsymbol{\pi}_k), \sigma_\tau(\boldsymbol{q}(\boldsymbol{\pi}_k)) - \boldsymbol{\pi}_k\big\rangle + \frac{4N\|\boldsymbol{R}\|_2^2}{\tau}\beta_k^2. \tag{SM2.3}$$

Next we modify the first-order terms by adding and subtracting $\sigma_\tau(\boldsymbol{q}(\boldsymbol{\pi}_k))$ inside the inner product; doing so yields

$$\begin{aligned}\big\langle \nabla_{\boldsymbol{\pi}}\mathcal{V}(\boldsymbol{\pi}_k), \sigma_\tau(\boldsymbol{q}(\boldsymbol{\pi}_k)) - \boldsymbol{\pi}_k\big\rangle &= \big\langle \nabla_{\boldsymbol{\pi}}\mathcal{V}(\boldsymbol{\pi}_k), \sigma_\tau(\boldsymbol{q}(\boldsymbol{\pi}_k)) - \boldsymbol{\pi}_k\big\rangle \\ &\quad + \big\langle \nabla_{\boldsymbol{\pi}}\mathcal{V}(\boldsymbol{\pi}_k), \sigma_\tau(\boldsymbol{q}(\boldsymbol{\pi}_k)) - \sigma_\tau(\boldsymbol{q}(\boldsymbol{\pi}_k))\big\rangle.\end{aligned} \tag{SM2.4}$$

Applying Lemma 4.1(b) and Lemma SM2.1(b) gives upper bounds on each of the terms in the RHS of equation (SM2.4)—namely

$$\big\langle \nabla_{\boldsymbol{\pi}}\mathcal{V}(\boldsymbol{\pi}_k), \sigma_\tau(\boldsymbol{q}(\boldsymbol{\pi}_k)) - \boldsymbol{\pi}_k\big\rangle \leq -\mathcal{V}\big(\boldsymbol{\pi}_k\big) + N\tau\log A_{\max}, \quad \text{and} \tag{SM2.5a}$$

$$\big\langle \nabla_{\boldsymbol{\pi}}\mathcal{V}(\boldsymbol{\pi}_k), \sigma_\tau(\boldsymbol{q}(\boldsymbol{\pi}_k)) - \sigma_\tau(\boldsymbol{q}(\boldsymbol{\pi}_k))\big\rangle \leq r\mathcal{V}(\boldsymbol{\pi}_k) + N\tau\log A_{\max} + \frac{\|\boldsymbol{R}\|_2^2}{2r\tau^3}\mathcal{W}(\boldsymbol{\pi}_k, \boldsymbol{q}_k). \tag{SM2.5b}$$

Combining equations (SM2.3) and (SM2.5) yields the claimed inequality (SM2.2a).

Proof of the $\mathcal{W}$-inequality (SM2.2b). Let us restate the TD update (2.8a) for convenience:

$$\boldsymbol{q}_{k+1} = \boldsymbol{q}_k + \alpha_k\,\frac{\boldsymbol{E}(\boldsymbol{a}_k)}{\boldsymbol{\pi}_k(\boldsymbol{a}_k)}\big(\boldsymbol{R}\;\boldsymbol{e}(\boldsymbol{a}_k) - \boldsymbol{q}_k\big) \qquad \text{for } k = 1, 2, \ldots.$$

Adding and subtracting terms on each side of this equation yields

$$\begin{aligned}\boldsymbol{q}_{k+1} - \boldsymbol{R}\boldsymbol{\pi}_{k+1} = \boldsymbol{q}_k - \boldsymbol{R}\boldsymbol{\pi}_k + \boldsymbol{R}(\boldsymbol{\pi}_k - \boldsymbol{\pi}_{k+1}) + \alpha_k\,\frac{\boldsymbol{E}(\boldsymbol{a}_k)}{\boldsymbol{\pi}_k(\boldsymbol{a}_k)}\big(\boldsymbol{R}\;\boldsymbol{e}(\boldsymbol{a}_k) - \boldsymbol{q}_k\big) \\ + \alpha_k\frac{\boldsymbol{E}(\boldsymbol{a}_k)}{\boldsymbol{\pi}(\boldsymbol{a}_k)}\boldsymbol{R}\boldsymbol{\pi}_k - \alpha_k\frac{\boldsymbol{E}(\boldsymbol{a}_k)}{\boldsymbol{\pi}(\boldsymbol{a}_k)}\boldsymbol{R}\boldsymbol{\pi}_k.\end{aligned}$$

After some rearranging, by combining the last three terms above, we find that

$$\begin{aligned}\boldsymbol{w}_{k+1} = (1-\alpha_k)\boldsymbol{w}_k + \alpha_k\big(\boldsymbol{I} - \frac{\boldsymbol{E}(\boldsymbol{a}_k)}{\boldsymbol{\pi}_k(\boldsymbol{a}_k)}\big)\boldsymbol{w}_k \\ + \alpha_k\frac{\boldsymbol{E}(\boldsymbol{a}_k)}{\boldsymbol{\pi}_k(\boldsymbol{a}_k)}\boldsymbol{R}(\boldsymbol{e}(\boldsymbol{a}_k) - \boldsymbol{\pi}_k) + \boldsymbol{R}(\boldsymbol{\pi}_k - \boldsymbol{\pi}_{k+1}),\end{aligned} \tag{SM2.6}$$

where $\boldsymbol{w}_k := \boldsymbol{q}_k - \boldsymbol{R}\boldsymbol{\pi}_k$ and $\boldsymbol{I}$ is a diagonal block matrix of the same dimensions as $\boldsymbol{R}$ with the i^{th} diagonal block being the identity matrix in $\mathbb{R}^{|\mathcal{A}^i|\times|\mathcal{A}^i|}$ for $i \in \mathcal{N}$. Denote by $\mathbb{E}_k$ the conditional expectation operator given $\boldsymbol{\pi}_k$ and $\boldsymbol{w}_k$. Note that the second and third terms in Equation (SM2.6) have zero mean conditioned on $\boldsymbol{\pi}_k$ and $\boldsymbol{w}_k$. As a result, when we take the mean conditioned on $\boldsymbol{\pi}_k$ and $\boldsymbol{w}_k$ of the square of the second norm on both sides, four out of six cross-terms on the right-hand side immediately vanish.

Moreover, the expected value of one additional cross-term is also zero; in particular, we have

$$E_k\Big\langle(\boldsymbol{I} - \frac{\boldsymbol{E}(\boldsymbol{a}_k)}{\boldsymbol{\pi}_k(\boldsymbol{a}_k)})\boldsymbol{w}_k, \frac{\boldsymbol{E}(\boldsymbol{a}_k)}{\boldsymbol{\pi}_k(\boldsymbol{a}_k)}\boldsymbol{R}(\boldsymbol{e}(\boldsymbol{a}_k) - \boldsymbol{\pi}_k)\Big\rangle$$
$$= E_k\Big\langle\frac{\boldsymbol{E}(\boldsymbol{a}_k)}{\boldsymbol{\pi}_k(\boldsymbol{a}_k)}^{\mathsf{T}}(\boldsymbol{I} - \frac{\boldsymbol{E}(\boldsymbol{a}_k)}{\boldsymbol{\pi}_k(\boldsymbol{a}_k)})\boldsymbol{w}_k, \boldsymbol{R}(\boldsymbol{e}(\boldsymbol{a}_k) - \boldsymbol{\pi}_k)\Big\rangle.$$

In the above inner product, as we are conditioning on $\boldsymbol{\pi}_k$ and $\boldsymbol{w}_k$, the i^{th} block of the first term only depends on player i's action a_k^i, while the i^{th} block of the second term does not depend on a_k^i as $R^{(i,i)} = 0$. As the actions played by each player are independent, the expected value of the inner product above is zero. Dropping all the vanishing cross-terms after taking the mean conditioned on $\boldsymbol{\pi}_k$ and $\boldsymbol{w}_k$ of the square of the second norm on both sides of Equation (SM2.6) leaves us with

$$\text{(SM2.7)}\quad \mathbb{E}_k[\|\boldsymbol{w}_{k+1}\|_2^2] \le (1-\alpha_k)^2\ \mathbb{E}_k\big[\|\boldsymbol{w}_k\|_2^2\big] + \alpha_k^2\ \mathbb{E}_k[\|(\boldsymbol{I} - \frac{\boldsymbol{E}(\boldsymbol{a}_k)}{\boldsymbol{\pi}_k(\boldsymbol{a}_k)})\boldsymbol{w}_k\|_2^2]$$
$$+ \|\boldsymbol{R}\|_2^2\ \mathbb{E}_k[\|\boldsymbol{\pi}_k - \boldsymbol{\pi}_{k+1}\|_2^2] + \alpha_k^2\ \mathbb{E}_k\big[\|\frac{\boldsymbol{E}(\boldsymbol{a}_k)}{\boldsymbol{\pi}_k(\boldsymbol{a}_k)}\boldsymbol{R}(\boldsymbol{e}(\boldsymbol{a}_k) - \boldsymbol{\pi}_k)\|_2^2\big]$$
$$+ (1-\alpha_k)\langle\boldsymbol{w}_k, \boldsymbol{R}(\boldsymbol{\pi}_k - \boldsymbol{\pi}_{k+1})\rangle.$$

Our next step is to examine each of these terms and upper bound them appropriately.

<u>Bounding the second term:</u> The submultiplicativity of the operator norm gives

$$\mathbb{E}_k\big[\|(\boldsymbol{I} - \frac{\boldsymbol{E}(\boldsymbol{a}_k)}{\boldsymbol{\pi}_k(\boldsymbol{a}_k)})\boldsymbol{w}_k\|_2^2\big] \le E_k\big[\|\!|(\boldsymbol{I} - \frac{\boldsymbol{E}(\boldsymbol{a}_k)}{\boldsymbol{\pi}_k(\boldsymbol{a}_k)})\|\!|_2^2\big]\,\|\boldsymbol{w}_k\|_2^2$$
$$\le E_k\big[\frac{1}{\min_{i\in\mathcal{N}} {\pi^i}_k(a_k^i)^2}\big]\ \ \|\boldsymbol{w}_k\|_2^2$$
$$\le \frac{A_{\max}N\|\boldsymbol{w}_k\|_2^2}{\delta}.$$

The final bound follows by upper bounding each term in the expectation and noting that there at most $A_{\max}N$ number of terms in the expectation.

<u>Bounding the third term:</u> We bound this term via Hölder's inequality, which leads to

$$\|\!|\boldsymbol{R}\|\!|_2^2\ \mathbb{E}_k[\|\boldsymbol{\pi}_k - \boldsymbol{\pi}_{k+1}\|_2^2] = \beta_k^2\|\!|\boldsymbol{R}\|\!|_2^2\|\sigma_\tau(\boldsymbol{q}_k) - \boldsymbol{\pi}_k\|_2^2$$
$$\le 4N\beta_k^2\|\!|\boldsymbol{R}\|\!|_2^2.$$

<u>Bounding the fourth term:</u> Note that $\|\boldsymbol{R}(\boldsymbol{e}(\boldsymbol{a}_k) - \boldsymbol{\pi}_k)\|_\infty \le 2N$ due to the assumption that each element of $\boldsymbol{R}$ lies in $[-1, 1]$. Using the submultiplicativity of the matrix

norm, we have

$$\begin{aligned}
\mathbb{E}_k\big[\|\frac{\boldsymbol{E}(\boldsymbol{a}_k)}{\boldsymbol{\pi}_k(\boldsymbol{a}_k)}\boldsymbol{R}(\boldsymbol{e}(\boldsymbol{a}_k)-\boldsymbol{\pi}_k)\|_2^2\big] &\leq \mathbb{E}_k\big[\|\frac{\boldsymbol{E}(\boldsymbol{a}_k)}{\boldsymbol{\pi}_k(\boldsymbol{a}_k)}\|_2^2\|\boldsymbol{R}\|_2^2\|\boldsymbol{e}(\boldsymbol{a}_k)-\boldsymbol{\pi}_k\|_2^2\big] \\
&\overset{(i)}{\leq} 4N\|\boldsymbol{R}\|_2^2\mathbb{E}_k\big[\|\frac{\boldsymbol{E}(\boldsymbol{a}_k)}{\boldsymbol{\pi}_k(\boldsymbol{a}_k)}\|_2^2\big] \\
&\overset{(ii)}{\leq} \frac{4N^2A_{\max}\|\boldsymbol{R}\|_2^2}{\delta}.
\end{aligned}$$

In inequality (i), we use Hölder's inequality the same way as in equation (4.3); inequality (ii) follows from the way we bounded the second term above.

Bounding the fifth term: Plugging in the strategy update equation (2.8b) and performing some algebra yields

$$\begin{aligned}
\langle \boldsymbol{w}_k, \boldsymbol{R}(\boldsymbol{\pi}_k-\boldsymbol{\pi}_{k+1})\rangle &= -\beta_k\langle \boldsymbol{w}_k, \boldsymbol{R}(\sigma_\tau(\boldsymbol{q}_k)-\boldsymbol{\pi}_k)\rangle \\
&= -\beta_k\langle \boldsymbol{w}_k, \boldsymbol{R}(\sigma_\tau(\boldsymbol{q}_k)-\sigma_\tau(\boldsymbol{R}\boldsymbol{\pi}_k))\rangle \\
&\qquad - \beta_k\langle \boldsymbol{w}_k, \boldsymbol{R}(\sigma_\tau(\boldsymbol{R}\boldsymbol{\pi}_k)-\boldsymbol{\pi}_k)\rangle.
\end{aligned}$$

We now use the Cauchy-Schwarz inequality to bound each of these terms. Using the fact that the function $u \mapsto \sigma_\tau(u)$ is $1/\tau$-Lipschitz gives

$$-\langle \boldsymbol{w}_k, \boldsymbol{R}(\sigma_\tau(\boldsymbol{q}_k)-\sigma_\tau(\boldsymbol{R}\boldsymbol{\pi}_k))\rangle \leq \frac{\|\boldsymbol{w}_k\|_2^2\|\boldsymbol{R}\|_2}{\tau}.$$

Again using the Cauchy-Schwarz inequality results in

$$-\langle \boldsymbol{w}_k, \boldsymbol{R}(\sigma_\tau(\boldsymbol{R}\boldsymbol{\pi}_k)-\boldsymbol{\pi}_k)\rangle \leq \|\boldsymbol{w}_k\|_2\|\boldsymbol{R}\|_2\|\sigma_\tau(\boldsymbol{R}\boldsymbol{\pi}_k)-\boldsymbol{\pi}_k\|_2.$$

Finally, applying Young's inequality for any $c > 0$ and parameter $r \in (0,1)$ and Lemma SM2.1(a) gives

$$\begin{aligned}
\langle \boldsymbol{w}_k, \boldsymbol{R}(\sigma_\tau(\boldsymbol{R}\boldsymbol{\pi}_k)-\boldsymbol{\pi}_k)\rangle &\leq \frac{1}{2cr}\|\boldsymbol{w}_k\|_2^2\|\boldsymbol{R}\|_2^2 + \frac{cr}{2}\|\sigma_\tau(\boldsymbol{R}\boldsymbol{\pi}_k)-\boldsymbol{\pi}_k\|_2^2 \\
&\leq \frac{1}{2cr}\|\boldsymbol{w}_k\|_2^2\|\boldsymbol{R}\|_2^2 + \frac{cr\mathcal{V}(\boldsymbol{\pi}_k)}{\tau}.
\end{aligned}$$

Choosing $c = \tau$ gives

$$\langle \boldsymbol{w}_k, \boldsymbol{R}(\boldsymbol{\pi}_k-\boldsymbol{\pi}_{k+1})\rangle \leq \frac{2\beta_k}{\tau r}\|\boldsymbol{w}_k\|_2^2\|\boldsymbol{R}\|_2^2 + \beta_k r\mathcal{V}(\boldsymbol{\pi}_k).$$

Combining the upper bounds: Putting together the pieces yields the claimed $\mathcal{W}$-inequality (SM2.2b)—that is

$$\begin{aligned}
\mathbb{E}W_{k+1} \leq (1-\alpha_k)^2\mathbb{E}W_k + \alpha_k^2\frac{A_{\max}N\mathbb{E}W_k}{\delta} + 4N\beta_k^2\|\boldsymbol{R}\|_2^2 + \alpha_k^2\frac{4N^2A_{\max}\|\boldsymbol{R}\|_2^2}{\delta} \\
+ (1-\alpha_k)\frac{2\beta_k}{\tau r}\|\boldsymbol{R}\|_2^2\mathbb{E}W_k + (1-\alpha_k)\beta_k r\mathbb{E}\mathcal{V}(\boldsymbol{\pi}_k).
\end{aligned}$$

SM2.3. Choice of parameters (4.8a) **and** (4.8b)**.** Note that this choice ensures the bound $c_{\alpha,\beta}(r,\tau) \leq r$. While the specification for τ is only relevant later, the choice for $\beta(\epsilon, r, \delta)$ ensures that $\alpha_k = \frac{\beta_k}{c_{\alpha,\beta}(r,\tau)}$ satisfies the upper bound $\alpha_k \leq \frac{c_{\alpha,\beta}\delta}{A_{\max}N}$. Then

$$1-\alpha_k+\alpha_k^2\tfrac{A_{\max}N}{\delta} \leq 1-\alpha_k(1-r).$$

The choice of the timescale separation constant also ensures that

$$\frac{c_{\alpha,\beta}(r,\tau)}{1-r} = \frac{1}{\frac{4\|\boldsymbol{R}\|_2^2}{r\tau^3}} \le \frac{1}{1-2r+\frac{3\|\boldsymbol{R}\|_2^2}{r\tau^3}},$$

and hence

$$-\alpha_k(1-r) + \beta_k\tfrac{3\|\boldsymbol{R}\|_2^2}{r\tau^3} = \beta_k\Big(-\frac{1-r}{c_{\alpha,\beta}(r,\tau)} + \frac{3\|\boldsymbol{R}\|_2^2}{r\,\tau^3}\Big) \le -\beta_k(1-2r),$$

enabling us to match the contraction terms for V_k and W_k.

SM2.4. Proof of Lemma 4.3. We begin with the drift inequality (4.9) on the total Lyapunov values $\{T_k\}_{k\ge 1}$. By solving this drift inequality, we obtain bounds on the expected Nash gap at time $K+1$ for a trajectory $\{\boldsymbol{\pi}_k\}_{k\ge1}$ that remains δ-good until time K.

LEMMA SM2.2. *Suppose that the iterates $\{\boldsymbol{\pi}_k\}_{k\ge1}$ are δ-good for times $k = 1,\dots,K$. Then the following properties hold:*

(a) For the constant stepsize $\beta_k \equiv \beta$, we have

$$\mathbb{E}\,\mathrm{NG}(\boldsymbol{\pi}_{K+1}) \le (1-\beta(1-2r))^K T_1 + \frac{2N}{1-2r}\tau\log A_{\max} + \frac{5N^2 A_{\max}\|\boldsymbol{R}\|_2^2\beta}{(1-2r)c^2_{\alpha,\beta}(r,\tau)\delta}.$$

(b) For the inverse polynomial stepsize $\beta_k = \frac{\beta}{(k+k_0)^\eta}$ for some exponent $\eta\in(0,1)$ and offset $k_0 = \lceil(\frac{2\eta}{\beta})^{1/(1-\eta)}\rceil$, we have

$$\mathbb{E}\,\mathrm{NG}(\boldsymbol{\pi}_{K+1}) \le \frac{\exp -\frac{(1-2r)\beta}{(1-\eta)}(K+k_0+1)^{1-\eta}}{\exp -\frac{(1-2r)\beta}{(1-\eta)}(1+k_0)^{\eta}} T_1 + \frac{2N}{1-2r}\tau\log A_{\max} + \frac{10N^2 A_{\max}\|\boldsymbol{R}\|_2^2\beta}{(1-2r)c^2_{\alpha,\beta}(r,\tau)\delta(K+k_0)^\eta}.$$

The results claimed in Lemma 4.3 then follow from Lemma SM2.2 by setting each term in the upper bounds on the Nash gap to be equal to $\frac{\epsilon}{3}$.

Proof of Lemma SM2.2. We split our analysis into different cases, depending on the choice of stepsizes.

Constant stepsizes. For constant stepsizes $\beta_k\equiv\beta\in(0,1)$, by iterating the drift inequality (4.9), we find that

$$\text{(SM2.8)}\quad \mathbb{E}T_{K+1} \le \big(1-\beta\big(1-2r\big)\big)^K\mathbb{E}T_1 + 2N\tau\log A_{\max}\sum_{j=1}^K \beta\big(1-(1-2r)\beta\big)^{K-j} + \frac{5A_{\max}N^2\|\boldsymbol{R}\|_2^2}{c^2_{\alpha,\beta}(r,\tau)\delta}\sum_{j=1}^K\beta^2\big(1-(1-2r)\beta\big)^{K-j}.$$

Bounding the second term: Consider the sequence $\{u_k\}_{k\ge1}$ defined with the initialization $u_1 = 0$ and the recursion $u_{k+1} = \big(1-(1-2r)\beta\big)u_k + (1-2r)\beta$. Then for

$K = 1, 2, \ldots$, we have

$$u_{K+1} = \sum_{j=1}^{K} (1-2r)\beta\big(1-(1-2r)\beta\big)^{K-j}.$$

Since $\beta \leq 1$ and $r < 0.5$, it follows from an inductive argument that $u_k \leq 1$ for all $k = 1, 2, \ldots$. Using this fact, we have

$$2N\tau \log A_{\max} \sum_{j=1}^{K} \beta\big(1-(1-2r)\beta\big)^{K-j} \leq \frac{2N}{1-2r}\tau \log A_{\max}.$$

Bounding the third term: Note that

$$\sum_{j=1}^{K} \big(1-(1-2r)\beta\big)^{K-j} = \sum_{j=0}^{K} \big(1-(1-2r)\beta\big)^{j} \leq \frac{1}{(1-2r)\beta}.$$

Using this fact, it follows that

$$\sum_{j=1}^{K} \beta^2\big(1-(1-2r)\beta\big)^{K-j} \leq \frac{\beta}{1-2r}.$$

Thus, equation (SM2.8) can be simplified to the form in Lemma SM2.2(a).

Inverse polynomial stepsizes. By assuming that the iterates $\boldsymbol{\pi}_k$ are δ-good for all $k = 0, 1, \ldots, K$, we have

$$\begin{aligned} T_{K+1} \leq \prod_{i=1}^{K} \big(1-(1-2r)\beta_i\big)T_1 + 2N\tau \log A_{\max} \sum_{i=1}^{K} \beta_i \prod_{j=i+1}^{K} \big(1-(1-2r)\beta_j\big) \\ + \frac{5N^2 A_{\max} \|\boldsymbol{R}\|_2^2}{c_{\alpha,\beta}^2(r,\tau)\delta} \sum_{i=1}^{K} \beta_i^2 \prod_{j=i+1}^{K} \big(1-(1-2r)\beta_j\big). \end{aligned}$$

Bounding the first term: In this case, we have

$$\begin{aligned} \prod_{i=1}^{K} \big(1-(1-2r)\beta_i\big) &= \exp\Big(\sum_{i=1}^{K} \log\big(1-(1-2r)\beta_i\big)\Big) \\ &\overset{(i)}{\leq} \exp\Big(-\sum_{i=1}^{K}(1-2r)\beta_i\Big) \\ &\overset{(ii)}{\leq} \exp\Big(-(1-2r)\beta \int_1^{K+1} \frac{1}{(x+k_0)^\eta} dx\Big) \\ &= \exp\Big(-\frac{(1-2r)\beta}{(1-\eta)}\big((K+k_0+1)^{1-\eta} - (1+k_0)^\eta\big)\Big). \end{aligned}$$

where step (i) follows from the inequality $\log(1-x) \leq -x$ for $x \in [0, 1)$, whereas step (ii) follows from bounding the Riemann sum.

Bounding the second term: Consider the sequence $\{u_k\}_{k\geq 1}$ defined by the initialization $u_1 = 0$ followed by the recursion $u_{k+1} = \big(1-(1-2r)\beta_k\big)u_k + (1-2r)\beta_k$. Then for

$K = 1, 2, \ldots$, we have

$$u_{K+1} = \sum_{i=1}^{K} (1-2r)\beta_i \prod_{j=i+1}^{K} \big(1-(1-2r)\beta_j\big).$$

Note that $\beta_k \leq 1$ for $k = 1, 2 \ldots$ and $r < \frac{1}{2}$ by assumption; therefore, it follows from an inductive argument that $u_k \leq 1$ for all $k = 1, 2, \ldots$. Using this fact, we have the bound

$$2N\tau \log A_{\max} \sum_{i=1}^{K} \beta_i \prod_{j=i+1}^{K} \big(1-(1-2r)\beta_j\big) \leq \frac{2N}{1-2r}\tau \log A_{\max}.$$

<u>Bounding the third term:</u> Consider the sequence $\{u_k\}_{k\geq 1}$ defined by the initialization $u_1 = 0$ and the recursion $u_{k+1} = \big(1-(1-2r)\beta_k\big)u_k + \beta_k^2$. Then for $K = 1, 2, \ldots$, we have

$$u_{K+1} = \sum_{i=1}^{K} \beta_i^2 \prod_{j=i+1}^{K} \big(1-(1-2r)\beta_j\big).$$

It can be shown by induction [SM1, see p.36] that $u_k \leq \frac{2}{(1-2r)}\beta_k$ for $k = 1, 2, \ldots$, from which it follows that

$$\frac{5N^2 A_{\max} \|\boldsymbol{R}\|_2^2}{c_{\alpha,\beta}^2(r,\tau)\delta} \sum_{i=1}^{K} \beta_i^2 \prod_{j=i+1}^{K} \big(1-(1-2r)\beta_j\big) \leq \frac{10N^2 A_{\max} \|\boldsymbol{R}\|_2^2 \beta}{(1-2r)c_{\alpha,\beta}^2(r,\tau)\delta(K+k_0)^\eta}. \qquad \square$$

SM2.5. Proof of Lemma 4.4. We prove the following more general claim:

LEMMA SM2.3. *Suppose that Algorithm 2.2 is initialized with uniform initial strategies $\boldsymbol{\pi}_1$ and the stepsize parameter initialized according to equation (4.8a). The quantity $\xi(\nu)$ denotes the ratio between the lower bound on $K_{good}(\beta,\delta)$ and $K(\epsilon,\delta,r)$.*

(a) For any $\xi(\nu) \in (1, 1+\nu)$, there exists a choice $r(\nu) \in (0, \frac{1}{2})$ such that Algorithm 2.2 with the constant stepsize $\beta_k \equiv \beta(\epsilon, r(\nu), \delta(\epsilon,\nu))$ and $\delta(\epsilon,\nu) := \big(\frac{\epsilon}{3T_1}\big)^{1+\nu}\frac{1}{A_{\max}}$ satisfies

$$K_{good}\big(\beta(\epsilon, r(\epsilon,\nu), \delta(\epsilon,\nu)), \delta(\epsilon,\nu)\big) \geq \frac{\xi(\nu)\log\big(\frac{\epsilon}{3T_1}\big)}{\log\big(1-(1-2r(\nu))\beta(\epsilon, r(\nu), \delta(\epsilon,\nu))\big)}.$$

(b) Consider the inverse polynomial stepsize $\beta_k = \frac{\beta(\epsilon,r,\delta(\epsilon,\nu))}{(k+k_0)^\eta}$ for some exponent $\eta \in (0,1)$, the offset $k_0 = \big\lceil (\frac{2\eta}{\beta})^{1/(1-\eta)} \big\rceil$, and $\delta(\epsilon,\nu) := \big(\frac{\epsilon}{3T_1}\big)^{1+\nu}\frac{e}{A_{\max}}$. Then for any choice of $\xi(\nu) \in (1, \sqrt{1+\nu})$ such that $\log(1-\beta_1) \geq -\xi(\nu)\beta_1$, there exists some $r(\nu) \in (0, \frac{1}{2})$ such that

$$K_{good}\big(\beta(\epsilon, r(\epsilon,\nu), \delta(\epsilon,\nu)), \delta(\epsilon,\nu)\big)$$
$$\geq \Big[\frac{\xi(\nu)}{1-2r}\frac{(1-\eta)}{\beta(\epsilon, r(\nu), \delta(\epsilon,\nu))}\log\big(\frac{3T_1}{\epsilon}\big) + (1+k_0)^{1-\eta}\Big]^{\frac{1}{1-\eta}} - k_0.$$

In the above lemma, the quantity $\xi(\nu) > 1$ is the factor by which $K_{\text{good}}(\beta,\delta)$ is greater than $K(\epsilon,\delta,r)$ (cf. Lemma 4.3), for the specific choices of δ and r mentioned in the lemma above. Our proof of Lemma SM2.3 exploits the following auxiliary result:

LEMMA SM2.4. *For any $\beta, \delta \in (0,1)$, either of the following two conditions are sufficient to ensure that the iterates $\{\boldsymbol{\pi}_k\}_{k\geq 1}$ of Algorithm 2.2 with timescale constant β are δ-good up to time K.*

(a) For the constant stepsize $\beta_k \equiv \beta$, it suffices to have $K \leq \frac{\log(A_{\max}\delta)}{\log(1-\beta)}$.

(b) For $\beta \in (0, \frac{1}{2})$, suppose that $\beta_k = \frac{\beta}{(k+k_0)^\eta}$ where $\eta \in (0,1)$, $k_0 = \big(\frac{2\eta}{\beta}\big)^{1/(1-\eta)}$, and $\xi > 1$ is chosen to ensure that $\log(1-\beta_1) \geq -\xi\beta_1$. Then it suffices to have

$$(K+k_0)^{1-\eta} \leq \frac{(1-\eta)}{\xi\beta} \log\big(\frac{e}{A_{\max}\delta}\big) + (1+k_0)^{1-\eta},$$

See Subsection SM3.2 for the proof. Equivalently, the upper bounds on K in Lemma SM2.4 are lower bounds on $K_{\text{good}}(\beta, \delta)$.

Now, we choose δ and r as a function of ϵ and ν to ensure that these lower bounds on $K_{\text{good}}(\beta,\delta)$ for the choice of β in Lemma 4.3 are bigger than $K(\epsilon,\delta,r)$. The statement in Lemma 4.4 follows for the specific choice of $\delta(\epsilon,\nu) = \big(\frac{\epsilon}{3T_1}\big)^{1+\nu}\frac{1}{A_{\max}}$ for the case of constant stepsizes and $\delta(\epsilon,\nu) = \big(\frac{\epsilon}{3T_1}\big)^{1+\nu}\frac{e}{A_{\max}}$ for the case of sublinearly decaying stepsizes. The exact value of r will be discussed below for each stepsize.

Constant stepsizes. In this case, we have $\beta_k \equiv \beta$ for any $\beta \in (0,1)$. Let $\xi \in (1, 1+\nu)$. There exists a small enough $r(\nu) \in (0, \frac{1}{2})$ such that

$$\frac{\log(1-\beta)}{\log\big(1-\beta\big(1-2r(\nu)\big)\big)} \leq \frac{1+\nu}{\xi}. \tag{SM2.10}$$

Applying this bound yields

$$\frac{\log(A_{\max}\ \delta(\epsilon,\nu))}{\log(1-\beta)} = (1+\nu)\frac{\log\big(\frac{\epsilon}{3T_1}\big)}{\log(1-\beta)} \geq \frac{\xi \log\big(\frac{\epsilon}{3T_1}\big)}{\log\big(1-\beta\big(1-2r(\nu)\big)\big)}.$$

Therefore, the iterates $\{\boldsymbol{\pi}_k\}_{k\geq 1}$ remain $\delta(\epsilon,\nu)$-good up until time

$$\frac{\xi \log\big(\frac{\epsilon}{3T_1}\big)}{\log\big(1-\beta\big(1-2r(\nu)\big)\big)}.$$

The statement of Lemma SM2.3 follows for $\beta = \beta(\epsilon, \delta(\epsilon,\nu), r(\nu))$.

Inverse polynomial stepsizes. Let $\xi(\nu) \in (1, \sqrt{1+\nu})$. If $\beta_k = \frac{\beta(\epsilon,\delta,r)}{(k+k_0)^\eta}$ where $\eta \in (0,1)$, $k_0 = \big(\frac{2\eta}{\beta}\big)^{1/(1-\eta)}$, as $\beta(\epsilon,\delta,r) \to 0$ as $r \to 0$ (equation (4.8a)), there will be a small enough $r(\nu)$ that satisfies $\log(1-\beta(\epsilon,\delta,r(\nu))) \geq -\xi(\nu)\beta(\epsilon,\delta,r(\nu))$. Find a small enough $r(\nu) \in (0,\frac{1}{2})$ such that we also have

$$1+\nu \geq \frac{\xi^2(\nu)}{1-2r(\nu)}. \tag{SM2.11}$$

Since $\xi^2(\nu) < 1+\nu$ by assumption, such a choice for $r(\nu)$ is possible. With such a choice, we have

$$\begin{aligned}\frac{1-\eta}{\xi(\nu)\beta(\epsilon,\delta(\epsilon,\nu),r(\nu))}\log\big(\frac{e}{A_{\max}\delta(\epsilon,\nu)}\big) &= (1+\nu)\frac{(1-\eta)}{\xi(\nu)\beta(\epsilon,\delta(\epsilon,\nu),r(\nu))}\log\big(\frac{3T_1}{\epsilon}\big)\\ &\geq \frac{\xi(\nu)(1-\eta)}{\beta(\epsilon,\delta(\epsilon,\nu),r(\nu))(1-2r(\nu))}\log\big(\frac{3T_1}{\epsilon}\big).\end{aligned}$$

Therefore, the iterates $\{\boldsymbol{\pi}_k\}_{k\geq 1}$ are $\delta(\epsilon,\nu)$-good for all K such that

$$(K+k_0)^{1-\eta} \leq \frac{\xi(\nu)}{1-2r(\nu)} \frac{(1-\eta)}{\beta(\epsilon,\delta(\epsilon,\nu),r(\nu))} \log\big(\frac{3T_1}{\epsilon}\big) + (1+k_0)^{1-\eta}.$$

SM3. Auxiliary results for Theorem 3.3. In this appendix, we collect together the proofs of various auxiliary results used in proving Theorem 3.3.

SM3.1. Proof of Lemma SM2.1. We prove each of the two parts in turn.

Proof of part (a). For any joint mixed strategy $\boldsymbol{\pi}$, define $\boldsymbol{\pi}^{-i}$ to be the collection of mixed strategies of all players except player i. Note that the average payoff function for player i, $q^i(\pi^{-i})$ is a function of $\boldsymbol{\pi}^{-i}$. Define the function

$$F^i(\pi^i,\boldsymbol{\pi}^{-i}) = \max_{\widehat{\pi}^i\in\Pi^i}\big\{\big(\widehat{\pi}^i-\pi^i\big)^\intercal q^i(\pi^{-i}) + \tau H(\widehat{\pi}^i) - \tau H(\pi^i)\big\}, \quad i=1,2,$$

and observe that $F^i(\boldsymbol{\pi})\geq 0$. Additionally, $F^i(\sigma_\tau(q^i(\pi^{-i})),\boldsymbol{\pi}^{-i}) = 0$. Therefore, by the first-order optimality condition and because the minimizer is in the relative interior of Π^i, we must have

$$\langle \nabla_{\pi^i} F^i(\sigma_\tau(q^i(\pi^{-i}))),\boldsymbol{\pi}^{-i}),\, \pi_1-\pi_2\rangle = 0 \quad \text{for all mixed strategies } \pi_1,\pi_2\in\Pi^i.$$

Note that as the negative of the Shannon entropy is 1-strongly convex on the probability simplex; consequently, the function F^i is τ-strongly convex with respect to π^i uniformly for all $\boldsymbol{\pi}^{-i}$. Therefore, we have

$$\begin{aligned}
F^i(\pi^i,\boldsymbol{\pi}^{-i}) &= F^i(\pi^i,\boldsymbol{\pi}^{-i}) - F^i(q^i(\pi^{-i})),\boldsymbol{\pi}^{-i})\\
&\geq \langle \nabla_{\pi^i} F^i(\sigma_\tau(q^i(\pi^{-i})),\boldsymbol{\pi}^{-i}),\, \pi^i-\sigma_\tau(q^i(\pi^{-i}))\rangle + \frac{\tau}{2}\big\|\sigma_\tau(q^i(\pi^{-i}))-\pi^i\big\|_2^2\\
&= \frac{\tau}{2}\big\|\sigma_\tau(q^i(\pi^{-i}))-\pi^i\big\|_2^2.
\end{aligned}$$

Using this lower bound, we can write

$$\frac{\tau}{2}\|\sigma_\tau(\boldsymbol{q}(\boldsymbol{\pi}))-\boldsymbol{\pi}\|_2^2 = \frac{\tau}{2}\sum_{i\in[N]}\big\|\sigma_\tau(q^i(\pi^{-i}))-\pi^i\big\|_2^2 \;\leq\; \sum_{i\in N} F^i(\pi^i,\boldsymbol{\pi}^{-i}) \overset{(i)}{\leq} \mathcal{V}(\boldsymbol{\pi}),$$

where step (i) follows from the non-negativity of the Shannon entropy. This completes the proof.

Proof of part (b). Let $\boldsymbol{v} = (v_i)_{i\in[N]}\in\otimes_{i\in[N]}\mathbb{R}^{|\mathcal{A}^i|}$. For each $i\in[N]$, consider the constrained optimization problem

$$\max_{\pi^i\in\mathbb{R}_+^{|\mathcal{A}_i|}}\big\{\langle\pi^i,\,v^i\rangle + \tau H(\pi^i)\big\} \qquad \text{such that } \textstyle\sum_{a\in\mathcal{A}^i}\pi^i(a)=1.$$

The gradient $\|\nabla H(\pi^i)\|_2$ diverges whenever any element of π^i approaches zero, so that we can argue that the optimum will be achieved at a vector ${\pi^i}_\star$ with strictly positive co-ordinates. We introduce a Lagrange multiplier λ for the sum-constraint. The associated KKT conditions imply the optimum ${\pi^i}_\star$ satisfies the condition

$$v^i + \tau\nabla H({\pi^i}_\star) + \lambda\boldsymbol{e} = 0,$$

where $\boldsymbol{e}$ is a vector of all ones, and λ is the Lagrange multiplier. Solving for the optimum and using the fact that λ is chosen to ensure that ${\pi^i}_\star$ satisfies the normalization constraint yields ${\pi^i}_\star = \sigma_\tau(v^i)$. Equivalently, we have shown that $v^i + \tau\nabla H(\sigma_\tau(v^i)) + \lambda \boldsymbol{e} = 0$, from which it follows that

$$\langle v^i + \tau\nabla H(\sigma_\tau(v^i)),\, \pi_1 - \pi_2\rangle = 0 \qquad \text{for any pair } \pi_1, \pi_2 \in \Pi^i.$$

The above result can also be expressed as

$$\langle \boldsymbol{v} + \tau\nabla H(\sigma_\tau(\boldsymbol{v})), \boldsymbol{\pi}_1 - \boldsymbol{\pi}_2\rangle = 0 \qquad \text{for any pair } \boldsymbol{\pi}_1, \boldsymbol{\pi}_2 \in \boldsymbol{\Pi}, \tag{SM3.1}$$

where $\nabla H(\sigma_\tau(\boldsymbol{v})) = (\nabla H(\sigma_\tau(v^i)))_{i\in[N]}$. We will now use this result to prove Lemma SM2.1(b).

Using the gradient we evaluated using Danskin's Theorem (SM2.1), we can write

$$\begin{aligned}
\big\langle \nabla_\pi \mathcal{V}(\boldsymbol{\pi}), \sigma_\tau(\boldsymbol{u}) - \sigma_\tau(\boldsymbol{q}(\boldsymbol{\pi}))\big\rangle &= \big\langle \boldsymbol{R}^\intercal \sigma_\tau(\boldsymbol{q}(\boldsymbol{\pi})), \sigma_\tau(\boldsymbol{u}) - \sigma_\tau(\boldsymbol{q}(\boldsymbol{\pi}))\big\rangle \\
&\overset{(i)}{=} \big\langle \sigma_\tau(\boldsymbol{q}(\boldsymbol{\pi})), \boldsymbol{R}\, \sigma_\tau(\boldsymbol{u})\big\rangle \\
&\overset{(ii)}{=} \big\langle \sigma_\tau(\boldsymbol{q}(\boldsymbol{\pi})) - \sigma_\tau(\boldsymbol{u}), \boldsymbol{R}\, \sigma_\tau(\boldsymbol{u})\big\rangle,
\end{aligned}$$

where steps (i) and (ii) follow from the zero-sum property (2.1). Next, choosing $\boldsymbol{v} = \boldsymbol{u}$ and applying the earlier result (SM3.1) gives

$$\begin{aligned}
\big\langle \sigma_\tau(\boldsymbol{q}(\boldsymbol{\pi})) - \sigma_\tau(\boldsymbol{u}), \boldsymbol{R}\, \sigma_\tau(\boldsymbol{u})\big\rangle &= \big\langle \sigma_\tau(\boldsymbol{u}) - \sigma_\tau(\boldsymbol{q}(\boldsymbol{\pi})), \tau\nabla H(\sigma_\tau(\boldsymbol{u}))\big\rangle \\
&\qquad + \big\langle \sigma_\tau(\boldsymbol{q}(\boldsymbol{\pi})) - \sigma_\tau(\boldsymbol{u}), \boldsymbol{R}\, \sigma_\tau(\boldsymbol{u}) - \boldsymbol{u}\big\rangle \\
&= S_1 + S_2,
\end{aligned}$$

where we define

$$\begin{aligned}
S_1 &:= \big\langle \sigma_\tau(\boldsymbol{u}) - \sigma_\tau(\boldsymbol{q}(\boldsymbol{\pi})), \tau\nabla H(\sigma_\tau(\boldsymbol{u}))\big\rangle, \quad \text{and} \\
S_2 &:= \big\langle \sigma_\tau(\boldsymbol{q}(\boldsymbol{\pi})) - \sigma_\tau(\boldsymbol{u}), \boldsymbol{R}\, \sigma_\tau(\boldsymbol{u}) - \boldsymbol{u}\big\rangle.
\end{aligned}$$

In order to complete the proof, it suffices to show that

$$S_1 \overset{(a)}{\leq} N\tau \log A_{\max}, \quad \text{and} \quad S_2 \overset{(b)}{\leq} r\mathcal{V}(\boldsymbol{\pi}) + \frac{\|\boldsymbol{R}\|_2^2}{2r\tau^3}\|\boldsymbol{q}(\boldsymbol{\pi}) - \boldsymbol{u}\|_2^2. \tag{SM3.2}$$

We prove each of these two claims in turn.

Proof of the bound (SM3.2)*(a).* Since H is a concave function, the first-order tangent bound implies that

$$\begin{aligned}
S_1 &= \sum_{i\in[N]} \big\langle \sigma_\tau(u^i) - \sigma_\tau(q^i(\pi^{-i})), \tau\nabla H(u^i)\big\rangle \\
&\leq \sum_{i\in[N]} \tau H(\sigma_\tau(u^i)) - \tau H(\sigma_\tau(q^i(\pi^{-i}))) \\
&\overset{(i)}{\leq} \sum_{i\in[N]} \tau H(\sigma_\tau(u^i)) \\
&\overset{(ii)}{\leq} N\tau \log A_{\max}.
\end{aligned}$$

In the above argument, step (i) follows from the non-negativity of H, whereas step (ii) follows from the fact that the discrete entropy is at most log cardinality of the space.

Proof of the bound (SM3.2)*(b).* We first begin by re-arranging as follows

$$
\begin{aligned}
S_2 &= \left\langle \sigma_\tau(\boldsymbol{q}(\boldsymbol{\pi})) - \sigma_\tau(\boldsymbol{u}), \boldsymbol{R}\ \sigma_\tau(\boldsymbol{u}) - \boldsymbol{u} \right\rangle \\
&= \left\langle \sigma_\tau(\boldsymbol{q}(\boldsymbol{\pi})) - \sigma_\tau(\boldsymbol{u}), \boldsymbol{R}\ \sigma_\tau(\boldsymbol{u}) - \boldsymbol{R}\ \sigma_\tau(\boldsymbol{R}\boldsymbol{\pi}) \right\rangle \\
&\qquad + \left\langle \sigma_\tau(\boldsymbol{q}(\boldsymbol{\pi})) - \sigma_\tau(\boldsymbol{u}), \boldsymbol{R}\ \sigma_\tau(\boldsymbol{R}\boldsymbol{\pi}) - \boldsymbol{R}\boldsymbol{\pi} \right\rangle + \left\langle \sigma_\tau(\boldsymbol{q}(\boldsymbol{\pi})) - \sigma_\tau(\boldsymbol{u}), \boldsymbol{R}\boldsymbol{\pi} - \boldsymbol{u} \right\rangle \\
&= \left\langle \sigma_\tau(\boldsymbol{q}(\boldsymbol{\pi})) - \sigma_\tau(\boldsymbol{u}), \boldsymbol{R}\ \sigma_\tau(\boldsymbol{R}\boldsymbol{\pi}) - \boldsymbol{R}\boldsymbol{\pi} \right\rangle + \left\langle \sigma_\tau(\boldsymbol{q}(\boldsymbol{\pi})) - \sigma_\tau(\boldsymbol{u}), \boldsymbol{R}\boldsymbol{\pi} - \boldsymbol{u} \right\rangle,
\end{aligned}
$$

where the final step follows from the zero-sum property (2.1). Using the fact that $\sigma_\tau(\cdot)$ is $1/\tau$-Lipschitz, the Cauchy-Schwarz inequality and the submultiplicativity of the matrix norm, we have

$$
S_2 \le \frac{1}{\tau} \|\boldsymbol{R}\|_2 \|\sigma_\tau(\boldsymbol{R}\boldsymbol{\pi}) - \boldsymbol{\pi}\|_2 \|\boldsymbol{q}(\boldsymbol{\pi}) - \boldsymbol{u}\|_2 + \frac{\|\boldsymbol{q}(\boldsymbol{\pi}) - \boldsymbol{u}\|_2^2}{\tau}.
$$

Applying the inequality $ab \le \frac{a^2 c + \frac{b^2}{c}}{2}$ with the choices $a = \|\sigma_\tau(\boldsymbol{R}\boldsymbol{\pi}) - \boldsymbol{\pi}\|_2$, $b = \frac{1}{\tau}\|\boldsymbol{R}\|_2 \|\boldsymbol{q}(\boldsymbol{\pi}) - \boldsymbol{u}\|_2$ and an arbitrary $c > 0$ combined with Lemma SM2.1(a) yields

$$
\begin{aligned}
S_2 &\le \frac{1}{2} c \|\sigma_\tau(\boldsymbol{R}\boldsymbol{\pi}) - \boldsymbol{\pi}\|_2^2 + \frac{\|\boldsymbol{R}\|_2^2}{2c\tau^2} \|\boldsymbol{q}(\boldsymbol{\pi}) - \boldsymbol{u}\|_2^2 \\
&\le \frac{c}{\tau} \mathcal{V}(\boldsymbol{\pi}) + \frac{\|\boldsymbol{R}\|_2^2}{2c\tau^2} \|\boldsymbol{q}(\boldsymbol{\pi}) - \boldsymbol{u}\|_2^2.
\end{aligned}
$$

Choosing our free parameter as $c = r\tau$ for some $r \in (0, 1)$, we find that

$$
S_2 \le r\mathcal{V}(\boldsymbol{\pi}) + \frac{\|\boldsymbol{R}\|_2^2}{2r\tau^3} \|\boldsymbol{q}(\boldsymbol{\pi}) - \boldsymbol{u}\|_2^2,
$$

as claimed.

SM3.2. Proof of Lemma SM2.4. From the dynamics (2.8b), we have the elementwise inequality ${\pi^i}_{k+1} \succeq {\pi^i}_k (1 - \beta_k)$. Iterating this inequality for $k = 1, \dots, K+1$ yields

$$
{\pi^i}_{K+1} \succeq \frac{1}{A_{\max}} \prod_{j=1}^{K} (1 - \beta_j).
$$

where we have made use of the lower bound $\boldsymbol{\pi}_1 \succeq \frac{1}{A_{\max}} \boldsymbol{e}$, as guaranteed by our uniform initialization.

Therefore, in order to ensure that ${\pi^i}_k \succeq \delta$ for all $k = 1, \dots, K+1$, it suffices to have

$$
\frac{1}{A_{\max}} \prod_{k=1}^{K} (1 - \beta_k) \ge \delta. \tag{SM3.3}
$$

Here we make use of the fact that $\beta_k \in (0, 1)$ for all k. We now analyze the condition (SM3.3) for our two different choices of stepsizes.

Constant stepsizes. For the constant stepsizes $\beta_k \equiv \beta \in (0, 1)$, condition (SM3.3) is equivalent to

$$
(1 - \beta)^K \ge A_{\max}\delta, \quad \text{or equivalently} \quad K \le \frac{\log(A_{\max}\delta)}{\log(1 - \beta)}.
$$

Inverse polynomial stepsizes. Now consider the inverse polynomial stepsizes $\beta_k = \frac{\beta}{(k+k_0)^\eta}$ for some exponent $\eta \in (0,1)$ and offset $k_0 = \big(\frac{2\eta}{\beta}\big)^{1/(1-\eta)}$. Choose $\xi > 1$ to be the smallest possible number such that $\log(1-\beta_1) \geq -\xi\beta_1$. For $\beta < \frac{1}{2}$, we also have $\xi\beta < 1$. Now, to find an elementwise lower bound for $\boldsymbol{\pi}_{K+1}$, we write

$$\begin{aligned}
\prod_{i=1}^{K}(1-\beta_i) &= \exp\big(\sum_{i=1}^{K}\log(1-\beta_i)\big) \\
&\overset{(i)}{\geq} \exp\big(-\sum_{i=1}^{K}\frac{\xi\beta}{(i+k_0)^\eta}\big) \\
&= \exp\big(-\frac{\xi\beta}{(1+k_0)^\eta}\big)\exp\big(-\sum_{i=2}^{K}\frac{\xi\beta}{(i+k_0)^\eta}\big) \\
&\overset{(ii)}{\geq} \exp\big(-\frac{\xi\beta}{(1+k_0)^\eta}\big)\exp\big(-\int_1^K\frac{\xi\beta}{(x+k_0)^\eta}dx\big) \\
&= \exp\big(-\frac{\xi\beta}{(1+k_0)^\eta}\big)\exp\big(-\frac{\xi\beta}{1-\eta}\big((K+k_0)^{1-\eta}-(1+k_0)^{1-\eta}\big)\big) \\
&\overset{(iii)}{\geq} \exp\big(-1\big)\exp\big(-\frac{\xi\beta}{1-\eta}\big((K+k_0)^{1-\eta}-(1+k_0)^{1-\eta}\big)\big),
\end{aligned}$$

where step (i) follows from the inequality $\log(1-\beta_k) \geq -\xi\beta_k$ for $k \in \mathbb{N}$; step (ii) follows from bounding the Riemann sum; and step (iii) follows from the fact that

$$\frac{\xi\beta}{(1+k_0)^\eta} \leq 1,$$

since $\xi\beta < 1$ for a small enough β. Now,

$$\begin{aligned}
&\exp\big(-1\big)\exp\big(-\frac{\xi\beta}{1-\eta}\big((K+k_0)^{1-\eta}-(1+k_0)^{1-\eta}\big)\big) \geq A_{\max}\delta \\
\implies\ &(K+k_0)^{1-\eta}-(1+k_0)^{1-\eta} \leq \frac{1-\eta}{\xi\beta}\log\big(\frac{e}{A_{\max}\delta}\big).
\end{aligned}$$

SM4. Analyzing $\|\boldsymbol{R}\|_2$ for a k-regular ring graph. Recall that the weighted adjacency matrix $\mathbf{A}$ is a circulant matrix for a k-regular graph. Let the first row of $\mathbf{A}$ be $(r_1,\ldots,r_N)$, where $r_m \in \{-1,0,1\}$ for $m \in \mathcal{N}$. We know that $r_1 = 0$ as the diagonal elements are zero for a weighted adjacency matrix. Since the graph $\mathcal{G}$ is k-regular, exactly k of the r_m's are non-zero. For the matrix $\mathbf{A}$ to be skew-symmetric, we also need the condition $r_m = -r_{N-m+2}$ for $m \in \mathcal{N}\setminus\{1\}$. Using the skew-symmetry, we have $\mathbf{A}^\top\mathbf{A} = -\mathbf{A}^2$, as a result of which $\|\mathbf{A}\|_2$ is equal to the largest absolute eigenvalue of $\mathbf{A}$. By standard Fourier analysis, the eigenvalues of a circulant matrix are of the form

$$\lambda_j = \sum_{m=1}^{N} r_m\omega_N^{j(m-1)}, \quad \text{for } j = 0,1,\ldots,N-1,$$

where $\omega_N = e^{2\pi i/N}$ is the N^{th} root of unity. The $(m-1)^{\text{th}}$ and $(N-m+1)^{\text{th}}$ powers of ω_N are symmetric around the real line, i.e., $e^{2\pi ij(m-1)N} = e^{-2\pi ij(N-m+1)/N}$. From the skew-symmetric property $r_m = -r_{N-m+2}$, for $m \leq \lceil N/2\rceil$ we have

$$r_m\omega_N^{j(m-1)} + r_{N-m+2}\omega_N^{j(N-m+1)} = 2ir_m\sin(2\pi j(m-1)/N).$$

It follows that the absolute value of any of the eigenvalues can be lower bounded as

$$|\lambda_j| \geq \Big| \sum_{1<m\leq \lceil \frac{N}{2} \rceil} r_m \sin(2\pi j(m-1)/N) \Big|.$$

This can be lower bounded further depending on whether $k/4$ is even or odd. Let

$$\lambda_{\text{lb}}^{(j)} := \lfloor \tfrac{k}{4} \rfloor \min_{1<m<\ell\leq \lceil \frac{N}{2} \rceil} |\sin(2\pi j(m-1)/N) - \sin(2\pi j(\ell-1)/N)|.$$

We then have the lower bound

$$|\lambda_j| \geq \begin{cases} \lambda_{\text{lb}}^{(j)}, & \frac{k}{4} \text{ even}, \\ \lambda_{\text{lb}}^{(j)} + \min_{1<m\leq \lceil \frac{N}{2} \rceil} |\sin(2\pi j(m-1)/N)|, & \frac{k}{4} \text{ odd}, \end{cases}$$

thereby showing that $\|\mathbf{A}\|_2 = \Omega(k)$. As we already have the bound $\|\mathbf{A}\|_2 \leq d_{\max}(\mathcal{G})$, it follows that $\|\boldsymbol{R}\|_2$ grows linearly with $d_{\max}(\mathcal{G})$ for this graph.

SM5. Consequences for two-player zero-sum games. Our results for general polymatrix games imply more specific guarantees for the canonical class of two-player zero-sum matrix games, which we describe here. We begin in Subsection SM5.1 with guarantees for the dynamics in Algorithm 2.1 that apply to the full information setting. In Subsection SM5.2, we turn to the two timescale updates in Algorithm 2.2 that apply to the minimal information setting.

A two-player zero-game is fully defined by the payoff matrix $\bar{R}$ for player 1, and payoff matrix $-\bar{R}^\top$ for player 2. Consequently, we can obtain upper bounds on the iteration complexity by setting $N = 2$ in the upper bounds from Theorems 3.1 and 3.3 and bounding the spectral norm $\|\boldsymbol{R}\|_2^2$ appropriately. In this very special case, we have $\boldsymbol{R} = \begin{bmatrix} 0 & \bar{R} \\ -\bar{R}^\top & 0 \end{bmatrix}$, from which it follows that $\|\boldsymbol{R}\|_2 = \|\bar{R}\|_2$.

SM5.1. Two-player zero-sum games with full information. We begin by stating finite-sample guarantees on the full information procedure summarized in Algorithm 2.1. The shorthand $V_1 := \mathcal{V}(\boldsymbol{\pi}_1)$ for the initial value of the Lyapunov function $\mathcal{V}$ from equation (2.9) used here corresponds to the case with $N = 2$.

COROLLARY SM5.1 (Nash gap finite-sample guarantees). *Consider the full information learning dynamics (Algorithm 2.1) initialized with $\tau = c_3\epsilon/A_{\max}$. Then the iteration complexity $K(\epsilon)$ is bounded as follows:*

(a) For the constant stepsizes $\beta_k \equiv \beta := \epsilon^2/(c_4\|\bar{R}\|_2^2 \log A_{\max})$,

$$K(\epsilon) \leq \frac{c_5\|\bar{R}\|_2^2 \log A_{\max}}{\epsilon^2} \log\big(\frac{V_1}{\epsilon}\big).$$

(b) For the inverse linear stepsize $\beta_k = \beta/k$ for some $\beta \in (1, 2]$,

$$K(\epsilon) \leq \frac{c_6\|\bar{R}\|_2^2 \beta^2 V_1 \log A_{\max}}{(\beta-1)\epsilon^2}.$$

(c) For the inverse polynomial stepsize $\beta_k = \beta/(k+k_0)^\eta$ where $\beta \in (0, 1)$, $\eta \in (0, 1)$ and $k_0 \geq \big(\frac{1-\eta}{\beta}\big)^{\frac{1}{1-\eta}}$,

$$K(\epsilon) \leq c_7(k_0^{1-\eta} + \frac{(1-\eta)}{\beta} \log \frac{V_1}{\epsilon}\big)^{1/(1-\eta)} + \big(\frac{c_8\|\bar{R}\|_2^2 \beta \log A_{\max}}{\epsilon^2}\big)^{1/\eta}.$$

The bounds on the iteration complexity in Corollary SM5.1 scales as $\mathcal{O}(\|\bar{R}\|_2^2/\epsilon^2)$.

SM5.2. Two-player zero-sum games with minimal information. We now turn to bounds applicable to Algorithm 2.2 that applies to the minimal information setting. In this case, we give explicit bounds on the *iteration complexity* $K(\epsilon)$, meaning the minimum number of rounds required to ensure that $\mathbb{E}\,\mathrm{NG}(\boldsymbol{\pi}_{K(\epsilon)+1}) \leq \epsilon$. We show the existence of a polynomial iteration complexity with a scaling of the order $(1/\epsilon)^{8+\nu}$, where the rate parameter $\nu > 0$ can be chosen arbitrarily close to zero. As was the case with zero-sum polymatrix games, the price of taking $\nu \to 0^+$ manifests in the growth of certain pre-factors; we use the notation $g(\nu)$ and variants thereof to indicate terms of this type.

Our result applies to Algorithm 2.2 where the temperature and the timescale separation constant are set as

$$\tau = \frac{g_\tau(\nu)\epsilon}{\log A_{\max}} \quad \text{and} \quad c_{\alpha,\beta} = \frac{g_{\alpha,\beta}(\nu)\tau^3}{\|\bar{R}\|_2^2} \quad \text{, respectively.} \tag{SM5.1}$$

Corollary SM5.2 (Nash gap finite-sample guarantees). *Suppose that Algorithm* 2.2 *is run with the parameters* (SM5.1), *and with the initial mixed strategies* $\boldsymbol{\pi}_1$ *being uniform. Then the iteration complexity* $K(\epsilon)$ *is bounded as follows:*

(a) For the constant stepsize $\beta_k \equiv \beta := \frac{g_1(\nu)\epsilon^{8+\nu}}{A_{\max}^7\|\bar{R}\|_2^6}$, *we have*

$$K(\epsilon) \leq K^\star(\epsilon,\nu) := \frac{A_{\max}^7\|\bar{R}\|_2^6}{g_2(\nu)\epsilon^{8+\nu}}\log\big(\frac{3T_1}{\epsilon}\big),$$

(b) For the inverse polynomial stepsize $\beta_k = \frac{\beta}{(k+k_0)^\eta}$ *for some exponent* $\eta \in (0,1)$, *offset* $k_0 = \lceil(\frac{2\eta}{\beta})^{1/(1-\eta)}\rceil$, *and* $\beta = \frac{g_1(\nu)\epsilon^{8+\nu}}{A_{\max}^7\|\bar{R}\|_2^6}$, *we have*

$$K(\epsilon) \leq K^\star(\epsilon,\nu) := \Big\{\frac{(1-\eta)A_{\max}^7\|\bar{R}\|_2^6}{g_3(\nu)\epsilon^{8+\nu}}\log\big(\frac{3T_1}{\epsilon}\big)\Big\}^{\frac{1}{1-\eta}}.$$

SM6. Connections among Lyapunov functions. It is worthwhile comparing the Lyapunov functions used in our analysis with those from related work. For tracking the strategy updates, the proofs of both Theorems 3.1 and 3.3 make use of the Lyapunov function

$$\mathcal{V}(\boldsymbol{\pi}) := \sum_{i=1}^N \max_{\hat{\pi}\in\Pi^i}\Big\{\hat{\pi}^\top q^i(\pi^{-i}) + \tau H(\hat{\pi})\Big\}, \tag{SM6.1a}$$

where H is the Shannon entropy (2.4b). (In addition, the proof of Theorem 3.3 also exploits an additional Lyapunov function for tracking the q-updates.)

As noted previously, our Lyapunov function $\mathcal{V}$ is an entropically-regularized N-player version of the Lyapunov function used by Harris [SM4] for two-player zero-sum matrix games. Using the notation of our paper, this N-player extension of Harris' function takes the form

$$\mathcal{V}_{\mathrm{H}}(\boldsymbol{\pi}) := \sum_{i=1}^N \max_{\widehat{\pi}\in\Pi^i} \hat{\pi}^\top q^i(\pi^{-i}). \tag{SM6.1b}$$

The work of Chen et al. [SM2] makes use of an alternative Lyapunov function for two-player zero-sum matrix games; its N-player extension is given by

$$\mathcal{V}_{\mathrm{alt}}(\boldsymbol{\pi}) := \sum_{i=1}^N \Big[\max_{\widehat{\pi}\in\Pi^i}\big\{\hat{\pi}^\top q^i(\pi^{-i}) + \tau H(\widehat{\pi})\big\} - \tau H(\pi^i)\Big]. \tag{SM6.1c}$$

Observe that this function differs from our Lyapunov function (SM6.1a) by the subtraction of the additional entropy terms (i.e., the term $-\tau H(\pi^i)$). As we discuss below, for zero-sum games, this Lyapunov function (SM6.1c) has a natural interpretation in terms of the Kullback–Leibler divergence.

Connections to $\mathcal{V}_{\mathbf{H}}$. Let us first compare and contrast $\mathcal{V}$ with the Harris Lyapunov function $\mathcal{V}_{\mathrm{H}}$ from equation (SM6.1b). The Lyapunov function $\mathcal{V}_{\mathrm{H}}$ is directly related to Nash equilibria, since we have $\mathcal{V}_{\mathrm{H}}(\boldsymbol{\pi}) = 0$ whenever $\boldsymbol{\pi}$ is a Nash equilibrium. In contrast, since the Shannon entropy is non-negative, our Lyapunov function (SM6.1a) need not be zero at a Nash equilibrium nor at a τ-regularized Nash equilibrium. Although we cannot hope to drive down $\mathcal{V}(\boldsymbol{\pi})$ to zero as $\boldsymbol{\pi}$ approaches a Nash equilibrium (as in a standard Lyapunov analysis), our function $\mathcal{V}$ does have three key properties that enable our analysis:

(a) First, from equation (2.10), we have the bound $\mathrm{NG}(\boldsymbol{\pi}) \leq \mathcal{V}(\boldsymbol{\pi})$, so that any strategy $\boldsymbol{\pi}$ with $\mathcal{V}(\boldsymbol{\pi}) \leq \epsilon$ has Nash gap at most ϵ.

(b) Second, we have the approximation error bound

$$\left|\mathcal{V}(\boldsymbol{\pi}) - \mathcal{V}_{\mathrm{H}}(\boldsymbol{\pi})\right| \leq N\tau \log A_{\max}, \tag{SM6.2a}$$

so that $\mathcal{V}$ is a good approximation to $\mathcal{V}_{\mathrm{H}}$ for small τ.

(c) Third, the Lyapunov function $\mathcal{V}$ is differentiable and $L := \frac{\|\boldsymbol{R}\|_2^2}{\tau}$-smooth with respect to the Euclidean norm on $\boldsymbol{\Pi}$, meaning that

$$\|\nabla\mathcal{V}(\boldsymbol{\pi}) - \nabla\mathcal{V}(\tilde{\boldsymbol{\pi}})\|_2 \leq L\|\boldsymbol{\pi} - \tilde{\boldsymbol{\pi}}\|_2 \qquad \text{for all } \boldsymbol{\pi}, \tilde{\boldsymbol{\pi}} \in \boldsymbol{\Pi}. \tag{SM6.2b}$$

Connections to $\mathcal{V}_{\mathbf{alt}}$. The $(1/\tau)$-scaling of the smoothness constant in item (c) is crucial to our analysis, and underlies our choice of Lypunov function $\mathcal{V}$, as opposed to the functions $\mathcal{V}_{\mathrm{H}}$ and $\mathcal{V}_{\mathrm{alt}}$ used in past work. Our analysis based on $\mathcal{V}$ enables us to prove a finite-sample guarantee with polynomial scaling in $(1/\epsilon)$. In contrast, for the Lyapunov function (SM6.1c) used by Chen et al. [SM2], the smoothness constant increases exponentially in the inverse temperature $(1/\tau)$. Since obtaining an ϵ-Nash-gap requires reducing the temperature, this scaling means that the finite-sample guarantees in the paper [SM2] also scale exponentially in $(1/\epsilon)$.

As a side-remark, we note that the Lyapunov function (SM6.1c)—despite its less desirable smoothness properties—has a very natural interpretation in terms of the Kullback–Leibler (KL) divergence $\mathrm{KL}(p, q)$ between two discrete distributions p and q. With this notation, consider the Lyapunov function

$$\mathcal{V}_{\mathrm{KL}}(\boldsymbol{\pi}) := \sum_{i=1}^{N} \mathrm{KL}(\pi^i, \sigma_\tau(q^i(\pi^{-i}))). \tag{SM6.3}$$

By standard properties of the KL divergence, it can be seen that this function is zero if and only if $\pi^i = \sigma_\tau(q^i(\pi^{-i}))$ for $i \in \mathcal{N}$, so its minima correspond to the set of τ-regularized Nash equilibria (2.5). In fact, for any zero-sum game, the KL-based function (SM6.3) is proportional to $\mathcal{V}_{\mathrm{alt}}$, and we can use it to show that τ-regularized Nash equilibria are unique. We summarize these facts in the following:

PROPOSITION SM6.1. *For any zero-sum polymatrix game and any $\tau > 0$, we have the equivalence*

$$\mathcal{V}_{alt}(\boldsymbol{\pi}) \equiv \tau\, \mathcal{V}_{KL}(\boldsymbol{\pi}). \tag{SM6.4}$$

Moreover, there is a unique τ-regularized Nash equilibrium (2.5)*, corresponding to the unique global minimum of the function $\mathcal{V}_{\text{KL}}$.*

See Subsection SM6.1 for the proof of this claim.

The proof of the equivalence (SM6.4) hinges crucially on the zero-sum nature of the polymatrix game. Since $\mathcal{V}_{\text{alt}}$ is a strictly convex function by inspection, it implies that the KL-based function $\mathcal{V}_{\text{KL}}$ is strictly convex for any zero-sum polymatrix game. (Again, this strict convexity need not be true in general.) The global minima of $\mathcal{V}_{\text{KL}}$ are equivalent to τ-regularized Nash equilibria, and since $\mathcal{V}_{\text{KL}}$ is strictly convex, the claimed uniqueness condition follows.

SM6.1. Proof of Proposition SM6.1. We prove the claimed equivalence in equation (SM6.4) as a consequence of the following auxiliary result: the KL-based function $\mathcal{V}_{\text{KL}}$ can be written as

$$\mathcal{V}_{\text{KL}}(\boldsymbol{\pi}) = \sum_{i=1}^{N} \big\{ -H(\pi^i) + \mathcal{G}_\tau(q^i(\pi^{-i}))\big\}, \tag{SM6.5}$$

where $\mathcal{G}_\tau(\theta) := \log\big(\sum_{a\in\mathcal{A}^i} e^{\theta(a)/\tau}\big)$ for any vector $\theta \in \mathbb{R}^{|\mathcal{A}^i|}$.

Taking this auxiliary result (SM6.5) as given, let us complete the proof of the equivalence (SM6.4). By standard results on exponential families (e.g., [SM5]), the function $\mathcal{G}_\tau$ is convex, and its Legendre dual (up to a rescaling by τ) is the Shannon entropy. Consequently, we can write

$$\mathcal{G}_\tau(q^i(\pi^{-i})) = \max_{\widehat{\pi}\in\Pi^i} \big\{ \big\langle \widehat{\pi}, \frac{1}{\tau} q^i(\pi^{-i})\big\rangle + H(\widehat{\pi})\big\} = \frac{1}{\tau}\max_{\widehat{\pi}\in\Pi^i}\big\{\big\langle \widehat{\pi}, q^i(\pi^{-i})\big\rangle + \tau H(\widehat{\pi})\big\}.$$

Substituting this equivalence into the auxiliary claim (SM6.5) and re-arranging yields the claimed equivalence $\mathcal{V}_{\text{alt}}(\boldsymbol{\pi}) \equiv \tau\mathcal{V}_{\text{KL}}(\boldsymbol{\pi})$.

It remains to prove the auxiliary claim (SM6.5). By definition of the function $\mathcal{V}_{\text{KL}}$, we have

$$\begin{aligned}
\mathcal{V}_{\text{KL}}(\boldsymbol{\pi}) &= \sum_{i=1}^{N} \text{KL}\left(\pi^i, \sigma_\tau\big(q^i(\pi^{-i})\big)\right) \\
&\overset{(i)}{=} \tau \sum_{i=1}^{N} \big\{ -H\big(\pi^i\big) - \big\langle \pi^i, \log\sigma_\tau\big(q^i(\pi^{-i})\big)\big\rangle\big\} \\
&\overset{(ii)}{=} \sum_{i=1}^{N} \big\{ -H\big(\pi^i\big) - \big\langle \pi^i, \frac{1}{\tau} q^i(\pi^{-i}) - G_\tau\big(q^i(\pi^{-i})\big)\mathbf{1}\big\rangle\big\} \\
&= \sum_{i=1}^{N} \big\{ -H\big(\pi^i\big) + \mathcal{G}_\tau\big(q^i(\pi^{-i})\big)\big\} - \frac{1}{\tau}\sum_{i=1}^{N}\big\langle \pi^i, q^i(\pi^{-i})\big\rangle \\
&\overset{(iii)}{=} \sum_{i=1}^{N} \big\{ -H\big(\pi^i\big) + G_\tau\big(q^i(\pi^{-i})\big)\big\},
\end{aligned}$$

where step (i) follows from the definition of the KL divergence; step (ii) follows from the definition of $\mathcal{G}_\tau$; and step (iii) follows from the zero-sum property.

Finally, we establish uniqueness of the τ-regularized Nash equilibrium (2.5). Note that any τ-regularized NE is a global minimizer of $\mathcal{V}_{\mathrm{KL}}$ over the set $\boldsymbol{\Delta}$. Since $\mathcal{V}_{\mathrm{KL}}$ is continuous and the set $\boldsymbol{\Delta}$ is compact, the global minimum is achieved. Since $\mathcal{V}_{\mathrm{alt}}$ is a strictly convex function, the equivalence (SM6.4) ensures that $\mathcal{V}_{\mathrm{KL}}$ is also strictly convex. Consequently, the function $\mathcal{V}_{\mathrm{KL}}$ has a unique global minimum, meaning that there is a unique τ-regularized NE.

REFERENCES

[SM1] Z. Chen, S. T. Maguluri, S. Shakkottai, and K. Shanmugam, *A Lyapunov theory for finite-sample guarantees of asynchronous Q-learning and TD-learning variants*, 2023, https://arxiv.org/abs/2102.01567.

[SM2] Z. Chen, K. Zhang, E. Mazumdar, A. Ozdaglar, and A. Wierman, *A finite-sample analysis of payoff-based independent learning in zero-sum stochastic games*, arXiv preprint arXiv:2303.03100, (2023).

[SM3] J. M. Danskin, *The theory of max-min and its application to weapons allocation problems*, Springer, 1967.

[SM4] C. Harris, *On the rate of convergence of continuous-time fictitious play*, Games and Economic Behavior, 22 (1998), pp. 238–259.

[SM5] M. J. Wainwright and M. I. Jordan, *Graphical models, exponential families and variational inference*, Foundations and Trends in Machine Learning, 1 (2008), pp. 1—305.