

Supplementary Material for: A New Local Rule for Convergence of ICLA to a Compatible Point

This supplementary material includes proofs of Lemmas 1 to 3 and Theorem 1 from Section IV.A. For the cited references in this material please see the references of the paper.

I. PROOF OF LEMMA 1

Proof: The proof for ICLA is somewhat similar to lemma 2 in [7]. Let's us define a mapping $T : \mathbf{K} \rightarrow \mathbf{K}$ as (I-1).

$$\forall i, \forall r \in \{1, \dots, m_i\} \quad \bar{p}_{ir} = \frac{p_{ir} + \phi_{ir}}{1 + \sum_{j=1}^{m_i} \phi_{ij}} \quad (\text{I-1})$$

Where $\psi_{ir} = d_{ir}(\underline{p}) - D_i(\underline{p})$ and $\phi_{ir}(\underline{p}) = \max\{\psi_{ir}(\underline{p}), 0\}$.

Because T is continuous and \mathbf{K} is closed, bounded and convex, therefore, based on Brouwer's fixed point theorem, T has at least one fixed point. We claim that a point will be a fixed point if and only if it is a compatible point of ICLA. To show this, assume that \underline{p} is a fixed point of T . Thus for all r and i :

$\bar{p}_{ir} = p_{ir}$. With respect to this fact and (I-1), we have:

$$\forall i, \forall r \in \{1, \dots, m_i\} \quad p_{ir} = \frac{p_{ir} + \phi_{ir}}{1 + \sum_{j=1}^{m_i} \phi_{ij}} \quad (\text{I-2})$$

From (I-2) we can conclude (I-3).

$$\sum_{j=1}^{m_i} \phi_{ij} = \frac{\phi_{ir}}{p_{ir}} \quad (\text{I-3})$$

Eq. (I-3) results in (I-4):

$$\frac{\phi_{i1}}{p_{i1}} = \frac{\phi_{i2}}{p_{i2}} = \dots = \frac{\phi_{ir}}{p_{ir}} \quad (\text{I-4})$$

The condition $\phi_{ir} > 0$ for all ϕ_{ir} is not possible, therefore, the only case that satisfies (I-4) is:

$$\forall i, m \quad \phi_{ir} = 0 \quad (\text{I-5})$$

Now assume that $y_i (1 \leq i \leq m_i)$ is an action that reaches maximum value for $d_{iy_i}(\underline{p})$ among the others. Consider another action, $y_j (j \neq i)$ that it's $d_{iy_j}(\underline{p})$ is smaller than $d_{iy_i}(\underline{p})$. Assume that probability distribution \underline{p} puts a positive probability value on y_j , thus we will have $D_i(\underline{p}) < d_{iy_j}(\underline{p})$, that means $\phi_{iy_j}(\underline{p}) > 0$. This is in contradiction with (I-5), therefore, the assumption is not valid and probability distribution \underline{p} puts zero probability values on all the actions that their $d_{iy_j}(\underline{p})$ are smaller than $d_{iy_i}(\underline{p})$. According to this fact, we have (I-6):

$$\forall \underline{q} \in \mathbf{K} \quad \sum_r d_{ir}(\underline{p}) \times p_{ir} \geq \sum_r d_{ir}(\underline{p}) \times q_{ir} \quad (\text{I-6})$$

Accordingly \underline{p} is a compatible point. Conversely if $\underline{p} \in \mathbf{K}$ is a compatible point, then (I-6) will be valid for all i . configuration \underline{q} also contains $\underline{q} = (p_1, \dots, e_r, \dots, p_n)$ for fixed i . Since $d_{ir}(\underline{p})$ is independent of p_i , we have $\psi_{ir}(\underline{p}) \leq 0$. Therefore $\phi_{ir}(\underline{p}) = 0$ for all i and all $r = 1, \dots, m_i$. This means \underline{p} is a fixed point of T . \square

II. PROOF OF LEMMA 2

Notation 1: \underline{p}_i^t is the probability vector of LA_i at iteration t and the other learning automata are not aware of its value.

Notation 2: $\hat{\underline{p}}_i^t$ is an estimation of \underline{p}_i^t by observing the chosen actions of LA_i during L .

Proof: When ICLA reaches a stable point, there exist a \underline{p}^* such that for every $\varepsilon > 0$:

$$\lim_{t \rightarrow \infty} P(|\underline{p}_i^t - \underline{p}_i^*| > \varepsilon) = 0 \quad (\text{II-1})$$

Obviously:

$$\lim_{t \rightarrow \infty} P(|\underline{p}_i^t - \underline{p}_i^*| > \varepsilon) = 0 \quad (\text{II-2})$$

Now if we can prove (II-3) the proof will be completed.

$$\lim_{t \rightarrow \infty} P(|\hat{p}_i^t - \underline{p}_i^*| > \varepsilon) = 0 \quad (\text{II-3})$$

In algorithm 1, $n_i^t(a_j)$ can be replaced by the summation of random variables X^t , where X^t is defined as:

$$X^t = \begin{cases} 1 & \text{if action } a_j \text{ was chosen during epoch t} \\ 0 & \text{otherwise} \end{cases}$$

Because changes of strategies in each epoch are negligible, therefore, the probability of choosing action a_j , during epoch t, is a fixed value such as \underline{p}_{ij}^t . Therefore:

$$E(X^t) = 1 \times \underline{p}_{ij}^t + 0 \times (1 - \underline{p}_{ij}^t) = \underline{p}_{ij}^t \quad (\text{II-4})$$

and

$$E(n_i^t(a_j)) = E\left(\frac{X_1^t + \dots + X_L^t}{L}\right) = \frac{1}{L} \times (E(X_1^t) + \dots + E(X_L^t)) = \frac{1}{L} \times L \times \underline{p}_{ij}^t = \underline{p}_{ij}^t \quad (\text{II-5})$$

Now, using algorithm 1 we can see that $\hat{p}_{ij}^{t+1} = \frac{1}{2} \times \underline{p}_{ij}^t + \frac{1}{2} \times \frac{n_i^t(a_j)}{L}$ and $\hat{p}_{ij}^t = \frac{1}{2^t} \times \underline{p}_{ij}^1 + \frac{1}{2^{t-1}} \times \frac{n_i^2(a_j)}{L} + \frac{1}{2^{t-2}} \times \frac{n_i^3(a_j)}{L} + \dots + \frac{1}{2} \times \frac{n_i^t(a_j)}{L}$, therefore, we have $E(\hat{p}_{ij}^{t+1}) = E\left(\frac{1}{2^t} \times \frac{n_i^1(a_j)}{L} + \dots + \frac{1}{2} \times \frac{n_i^t(a_j)}{L}\right)$ which can be rewritten as $E(\hat{p}_{ij}^{t+1}) = \frac{1}{2^t} \times \underline{p}_{ij}^1 + \dots + \frac{1}{2} \times \underline{p}_{ij}^t$. As a consequence we have (II-6):

$$\begin{aligned} \lim_{t \rightarrow \infty} E(\hat{p}_{ij}^{t+1}) &= \lim_{t \rightarrow \infty} \left(\frac{1}{2^t} \times \underline{p}_{ij}^1 + \dots + \frac{1}{2} \times \underline{p}_{ij}^t \right) = \\ &\dots + 0 + 0 + \dots + \frac{1}{2^k} \times \lim_{t \rightarrow \infty} (\underline{p}_{ij}^{t-k+1}) + \dots + \frac{1}{2} \times \lim_{t \rightarrow \infty} (\underline{p}_{ij}^t) \end{aligned} \quad (\text{II-6})$$

$$\begin{aligned} \lim_{t \rightarrow \infty} E(\hat{p}_{ij}^{t+1}) &= 0 + 0 + \dots + \frac{1}{2^2} \times \underline{p}_{ij}^* + \frac{1}{2} \times \underline{p}_{ij}^* = \\ &\underline{p}_{ij}^* \left(\sum_{t=1}^{\infty} \frac{1}{2^t} \right) = \underline{p}_{ij}^* \end{aligned} \quad (\text{II-7})$$

Therefore in general, we have:

$$\lim_{t \rightarrow \infty} E(\hat{p}_i^t) = \underline{p}_i^* \quad (\text{II-8})$$

Since \underline{p}_i^* is the expected value of \hat{p}_i^t , from Central limit theorem, we get:

$$\lim_{t \rightarrow \infty} P(|\hat{p}_i^t - \underline{p}_i^*| > \varepsilon) = 0 \quad (\text{II-9})$$

Thus the proof is completed. \square

III. PROOF OF LEMMA 3

Proof: Let to define a mapping $P_t : \{Q\} \rightarrow \{Q\}$ as (III-1).

$$\begin{aligned} P_t^i Q_t^i(a^1, \dots, a^{\bar{m}}) &= Er_i^t(a^1, \dots, a^{\bar{m}}) + \beta \times \\ &\sum_{a^1} \sum_{a^2} \dots \sum_{a^{\bar{m}}} \underline{p}_i^{br}(\hat{p}^t, Q_t^i)(a^i) \prod_{\substack{j=1 \\ j \neq i}}^{\bar{m}} \hat{p}_j^t(a^j) Q_t^i(a^1, \dots, a^{\bar{m}}) \end{aligned} \quad (\text{III-1})$$

We have also (III-2).

$$\begin{aligned} P_t^i Q_*^i(a^1, \dots, a^{\bar{m}}) &= Er_i^t(a^1, \dots, a^{\bar{m}}) + \beta \times \\ &\sum_{a^1} \sum_{a^2} \dots \sum_{a^{\bar{m}}} \underline{p}_i^{br}(\underline{p}^*, Q_*^i)(a^i) \prod_{\substack{j=1 \\ j \neq i}}^{\bar{m}} \underline{p}_j^*(a^j) Q_*^i(a^1, \dots, a^{\bar{m}}) \end{aligned} \quad (\text{III-2})$$

It's straight forward that

$$\begin{aligned} \|P_t^i Q_t^i(a^1, \dots, a^{\bar{m}}) - P_t^i Q_*^i(a^1, \dots, a^{\bar{m}})\| &= \|Er_i^t(a^1, \dots, a^{\bar{m}}) + \beta \times \\ &\sum_{a^1} \sum_{a^2} \dots \sum_{a^{\bar{m}}} \underline{p}_i^{br}(\hat{p}^t, Q_t^i)(a^i) \prod_{\substack{j=1 \\ j \neq i}}^{\bar{m}} \hat{p}_j^t(a^j) Q_t^i(a^1, \dots, a^{\bar{m}}) - \\ &Er_i^t(a^1, \dots, a^{\bar{m}}) - \beta \times \\ &\sum_{a^1} \sum_{a^2} \dots \sum_{a^{\bar{m}}} \underline{p}_i^{br}(\underline{p}^*, Q_*^i)(a^i) \prod_{\substack{j=1 \\ j \neq i}}^{\bar{m}} \underline{p}_j^*(a^j) Q_*^i(a^1, \dots, a^{\bar{m}})\| \end{aligned} \quad (\text{III-3})$$

Since result of (III-3) is independent of $Er_i^t(a^1, \dots, a^{\bar{m}})$, therefore, we have (III-4).

$$\| P_t^i Q_t^i(a^1, \dots, a^{\bar{m}}) - P_t^i Q_*^i(a^1, \dots, a^{\bar{m}}) \| = \beta \times \quad (\text{III-4})$$

$$\begin{aligned} & | \sum_{a^1} \sum_{a^2} \dots \sum_{a^{\bar{m}}} \underline{p}_i^{br}(\underline{p}^*, Q_t^i)(a^i) \prod_{\substack{j=1 \\ j \neq i}}^{\bar{m}} \hat{p}_j^t(a^j) Q_t^i(a^1, \dots, a^{\bar{m}}) - \\ & \sum_{a^1} \sum_{a^2} \dots \sum_{a^{\bar{m}}} \underline{p}_i^{br}(\underline{p}^*, Q_*^i)(a^i) \prod_{\substack{j=1 \\ j \neq i}}^{\bar{m}} \underline{p}_j^*(a^j) Q_*^i(a^1, \dots, a^{\bar{m}}) | \end{aligned}$$

Let to put $\hat{p}_i^t(a_j) = \underline{p}_i^*(a_j) + \varepsilon_i^t(a_j)$, therefore, we have
(III-5):

$$\begin{aligned} & \beta \cdot | \sum_{a^1} \sum_{a^2} \dots \sum_{a^{\bar{m}}} \underline{p}_i^{br}(\underline{p}^*, Q_t^i)(a^i) \prod_{\substack{j=1 \\ j \neq i}}^{\bar{m}} \hat{p}_j^t(a^j) Q_t^i(a^1, \dots, a^{\bar{m}}) - \\ & \sum_{a^1} \sum_{a^2} \dots \sum_{a^{\bar{m}}} \underline{p}_i^{br}(\underline{p}^*, Q_*^i)(a^i) \prod_{\substack{j=1 \\ j \neq i}}^{\bar{m}} \underline{p}_j^*(a^j) Q_*^i(a^1, \dots, a^{\bar{m}}) | \leq (\text{III-5}) \\ & \beta \cdot | \sum_{a^1} \sum_{a^2} \dots \sum_{a^{\bar{m}}} \underline{p}_i^{br}(\underline{p}^*, Q_t^i)(a^i) \prod_{\substack{j=1 \\ j \neq i}}^{\bar{m}} \underline{p}_j^*(a^j) Q_t^i(a^1, \dots, a^{\bar{m}}) - \\ & \sum_{a^1} \sum_{a^2} \dots \sum_{a^{\bar{m}}} \underline{p}_i^{br}(\underline{p}^*, Q_*^i)(a^i) \prod_{\substack{j=1 \\ j \neq i}}^{\bar{m}} \underline{p}_j^*(a^j) Q_*^i(a^1, \dots, a^{\bar{m}}) | + \\ & \beta \cdot | \sum_{a^1} \sum_{a^2} \dots \sum_{a^{\bar{m}}} \underline{p}_i^{br}(\underline{p}^*, Q_t^i)(a^i) \prod_{\substack{j=1 \\ j \neq i}}^{\bar{m}} \varepsilon_j^t(a^j) Q_t^i(a^1, \dots, a^{\bar{m}}) | \end{aligned}$$

Using (II-9), we have:

$$\lim_{t \rightarrow \infty} | \sum_{a^1} \dots \sum_{a^{\bar{m}}} \underline{p}_i^{br}(\underline{p}^*, Q_t^i)(a^i) \prod_{\substack{j=1 \\ j \neq i}}^{\bar{m}} \varepsilon_j^t(a^j) Q_t^i(a^1, \dots, a^{\bar{m}}) | \quad (\text{III-6})$$

$$= 0$$

Now let the following notations:

$$EQ_t = \sum_{a^1} \dots \sum_{a^{\bar{m}}} \underline{p}_i^{br}(\underline{p}^*, Q_t^i)(a^i) \prod_{\substack{j=1 \\ j \neq i}}^{\bar{m}} \underline{p}_j^*(a^j) Q_t^i(a^1, \dots, a^{\bar{m}}) \quad (\text{III-7})$$

$$EQ_* = \sum_{a^1} \dots \sum_{a^{\bar{m}}} \underline{p}_i^{br}(\underline{p}^*, Q_*^i)(a^i) \prod_{\substack{j=1 \\ j \neq i}}^{\bar{m}} \underline{p}_j^*(a^j) Q_*^i(a^1, \dots, a^{\bar{m}}) \quad (\text{III-8})$$

$$\lambda_t = \beta \times | \sum_{a^1} \dots \sum_{a^{\bar{m}}} \underline{p}_i^{br}(\underline{p}^*, Q_t^i)(a^i) \prod_{\substack{j=1 \\ j \neq i}}^{\bar{m}} \varepsilon_j^t(a^j) Q_t^i(a^1, \dots, a^{\bar{m}}) | \quad (\text{III-9})$$

Using (III-4) and (III-5) we have (III-10).

$$\| P_t^i Q_t^i(a^1, \dots, a^{\bar{m}}) - P_t^i Q_*^i(a^1, \dots, a^{\bar{m}}) \| \leq \beta \cdot | EQ_t - EQ_* | + \lambda_t \quad (\text{III-10})$$

From (II-9), we have $\underline{p}_i^{br}(\underline{p}^*, Q_t^i) \rightarrow \underline{p}_i^*(a^i)$ and $\underline{p}_i^{br}(\underline{p}^*, Q_*^i) \rightarrow \underline{p}_i^*(a^i)$ when $t \rightarrow \infty$. Now if $EQ_t \geq EQ_*$ then we have:

$$\begin{aligned} & | EQ_t - EQ_* | = EQ_t - EQ_* = \\ & \sum_{a^1} \dots \sum_{a^{\bar{m}}} \underline{p}_i^*(a^i) \prod_{\substack{j=1 \\ j \neq i}}^{\bar{m}} \underline{p}_j^*(a^j) Q_t^i(a^1, \dots, a^{\bar{m}}) - \\ & \sum_{a^1} \dots \sum_{a^{\bar{m}}} \underline{p}_i^*(a^i) \prod_{\substack{j=1 \\ j \neq i}}^{\bar{m}} \underline{p}_j^*(a^j) Q_*^i(a^1, \dots, a^{\bar{m}}) = \\ & \sum_{a^1} \dots \sum_{a^{\bar{m}}} \prod_{j=1}^{\bar{m}} \underline{p}_j^*(a^j) (Q_t^i(a^1, \dots, a^{\bar{m}}) - Q_*^i(a^1, \dots, a^{\bar{m}})) \leq \\ & \sum_{a^1} \dots \sum_{a^{\bar{m}}} \prod_{j=1}^{\bar{m}} \underline{p}_j^*(a^j) \times \max_{(a^1, \dots, a^{\bar{m}})} (Q_t^i(a^1, \dots, a^{\bar{m}}) - Q_*^i(a^1, \dots, a^{\bar{m}})) = \\ & 1 \times \max_{(a^1, \dots, a^{\bar{m}})} (Q_t^i(a^1, \dots, a^{\bar{m}}) - Q_*^i(a^1, \dots, a^{\bar{m}})) \leq \\ & \max_{(a^1, \dots, a^{\bar{m}})} | Q_t^i(a^1, \dots, a^{\bar{m}}) - Q_*^i(a^1, \dots, a^{\bar{m}}) | = \\ & \| Q_t^i(a^1, \dots, a^{\bar{m}}) - Q_*^i(a^1, \dots, a^{\bar{m}}) \| \end{aligned} \quad (\text{III-11})$$

Therefore, using (III-10) and (III-11) we have:

$$\begin{aligned} & \| P_t^i Q_t^i(a^1, \dots, a^{\bar{m}}) - P_t^i Q_*^i(a^1, \dots, a^{\bar{m}}) \| \leq \\ & \beta \cdot \| Q_t^i(a^1, \dots, a^{\bar{m}}) - Q_*^i(a^1, \dots, a^{\bar{m}}) \| + \lambda_t \end{aligned} \quad (\text{III-12})$$

If we have the other case ($EQ_t < EQ_*$) then:

$$\begin{aligned} & | EQ_t - EQ_* | = EQ_* - EQ_t = \\ & \sum_{a^1} \dots \sum_{a^{\bar{m}}} \prod_{j=1}^{\bar{m}} \underline{p}_j^*(a^j) (Q_*^i(a^1, \dots, a^{\bar{m}}) - Q_t^i(a^1, \dots, a^{\bar{m}})) \leq \\ & \sum_{a^1} \dots \sum_{a^{\bar{m}}} \prod_{j=1}^{\bar{m}} \underline{p}_j^*(a^j) \times \max_{(a^1, \dots, a^{\bar{m}})} (Q_*^i(a^1, \dots, a^{\bar{m}}) - Q_t^i(a^1, \dots, a^{\bar{m}})) \\ & \leq \max_{(a^1, \dots, a^{\bar{m}})} | (Q_*^i(a^1, \dots, a^{\bar{m}}) - Q_t^i(a^1, \dots, a^{\bar{m}})) | = \\ & \max_{(a^1, \dots, a^{\bar{m}})} | Q_t^i(a^1, \dots, a^{\bar{m}}) - Q_*^i(a^1, \dots, a^{\bar{m}}) | = \\ & \| Q_t^i(a^1, \dots, a^{\bar{m}}) - Q_*^i(a^1, \dots, a^{\bar{m}}) \| \end{aligned} \quad (\text{III-13})$$

Therefore, (III-12) is valid for the case $EQ_t < EQ_*$ as well. This result in (based on corollary 1, (II-9) and (III-12)) that updated Q-values by (8) results in convergence to optimal Q-values of (7), with probability one and the proof is completed. \square

IV. PROOF OF THEOREM 1

Proof: In lemma 3, we proved that updating Q-values using (8) causes convergence to optimal Q-values of (7) with probability one. Therefore, when $t \rightarrow \infty$, (8) will approach to (IV-1).

$$\begin{aligned} Q_*^i(a^1, \dots, a^{\bar{m}}) = & (1 - \alpha) \times Q_*^i(a^1, \dots, a^{\bar{m}}) + \alpha \times [Er_i^t(a^1, \dots, a^{\bar{m}}) + \\ & \beta \cdot \sum_{a^1} \sum_{a^2} \dots \sum_{a^{\bar{m}}} p_i^{br}(p^*, Q_*^i)(a^i) \prod_{\substack{j=1 \\ j \neq i}}^{\bar{m}} p_j^*(a^j) \cdot Q_*^i(a^1, \dots, a^{\bar{m}})] \end{aligned} \quad (\text{IV-1})$$

It's obvious that

$$\begin{aligned} Q_*^i(a^1, \dots, a^{\bar{m}}) = & Er_i^t(a^1, \dots, a^{\bar{m}}) + \\ & \beta \cdot \sum_{a^1} \sum_{a^2} \dots \sum_{a^{\bar{m}}} p_i^{br}(p^*, Q_*^i)(a^i) \prod_{\substack{j=1 \\ j \neq i}}^{\bar{m}} p_j^*(a^j) \cdot Q_*^i(a^1, \dots, a^{\bar{m}}) \end{aligned} \quad (\text{IV-2})$$

Comparing (7) and (IV-2) results in that

$$\begin{aligned} v^i(p_1^*, p_2^*, \dots, p_i^{br}(p^*, Q_*^i), \dots, p_{\bar{m}}^*) = & \\ \sum_{a^1} \sum_{a^2} \dots \sum_{a^{\bar{m}}} & p_i^{br}(p^*, Q_*^i)(a^i) \prod_{\substack{j=1 \\ j \neq i}}^{\bar{m}} p_j^*(a^j) \cdot Q_*^i(a^1, \dots, a^{\bar{m}}) \end{aligned} \quad (\text{IV-3})$$

In algorithm 2, we specify the best response strategy using RHS of (IV-3) which is equal to $p_i^{br}(p^*, Q_*^i)$ in $v^i(p_1^*, p_2^*, \dots, p_i^{br}(p^*, Q_*^i), \dots, p_{\bar{m}}^*)$. Therefore, algorithm 2 specifies a best response strategy to $(p_1^*, \dots, p_{i-1}^*, p_{i+1}^*, \dots, p_{\bar{m}_i}^*)$.

Based on learning algorithm L_{R-I} [24], this local rule leads the probability vector of LA to the best response configuration. Since every LA in ICLA applies this rule, therefore probability vectors of all LAs are best response configuration. Therefore, according to best response strategy playing concepts [20], ICLA converges to a compatible point.

□