the experiment ($\boldsymbol{\delta}$ and $\boldsymbol{\epsilon}$) and approximation ($\boldsymbol{\delta}'$ and $\boldsymbol{\epsilon}'$) is evaluated by

$$\text{error} := \frac{1}{4}\sum_{i=1}^{4}\sqrt{|\delta_i - \delta_i'|^2 + |\epsilon_i - \epsilon_i'|^2}. \tag{16}$$

5.2 Chaos-Like and Heteroclinic Dynamics

Interestingly, learning dynamics in multi-memory games are complex. Fig. 3 shows two learning dynamics between which there is a slight difference in their initial strategies ($\boldsymbol{x} = \boldsymbol{y} = 0.8 \times \mathbf{1}$ in the solid line, but in the broken line ($\boldsymbol{x}'$ and $\boldsymbol{y}'$), $x_1' = 0.801$ and others are the same as the solid line). We use Algorithm 2 with $\eta = 10^{-3}$ and $\gamma = 10^{-6}$. These dynamics are similar in the beginning ($0 \leq t \leq 320$). However, the difference between these dynamics is gradually amplified ($320 \leq t \leq 360$), leading to the crucial difference eventually ($360 \leq t \leq 420$). We here introduce the distance between $\boldsymbol{x}'$ and $\boldsymbol{x}$ as

$$D(\boldsymbol{x}', \boldsymbol{x}) := \frac{1}{4}\sum_{i=1}^{4}|L(x_i') - L(x_i)|, \tag{17}$$

with $L(x) := \log x - \log(1-x)$; $L(x)$ is the measure taking into account the weight in replicator dynamics. Furthermore, in order to analyze how the difference is amplified, Fig. 3 also shows the maximum eigenvalue in learning dynamics. We can see that the larger the maximum eigenvalue is, the more the difference between the two trajectories is amplified. We observe that such an amplification typically occurs when strategies are close to the boundary of the simplex. In conclusion, the learning dynamics provide chaos-like sensitivity to the initial condition.

5.3 Divergence in General Memories and Actions

Although we have focused on the one-memory two-action zero-sum games so far, numerical simulations demonstrate that similar phenomena are seen in games of other numbers of memories and actions. Fig. 4 shows the trajectories of learning dynamics in various multi-memory and multi-action games, where we use Algorithm 2 with $\eta = 10^{-2}$ and $\gamma = 10^{-6}$. Note that we consider zero-sum games in all the panels (see Fig. 4-A for the payoff matrices). In Fig. 4-B, each panel shows that strategy variables $x^{a|s}$ roughly diverge from the Nash equilibrium and sojourn longer at the edges of the simplex, i.e., $x^{a|s} = 0$ or 1. Furthermore, Kullback-Leibler divergence from the Nash equilibrium averaged over the whole states, i.e.,

$$D_{\mathrm{KL}}(\mathbf{x}^*\|\mathbf{x}) := \frac{1}{|\mathcal{S}|}\sum_{s\in\mathcal{S}}\sum_{a\in\mathcal{A}} x^{*a|s}\log\frac{x^{*a|s}}{x^{a|s}}. \tag{18}$$

also increases with time in each panel of the figure. Thus, we confirm that learning reaches heteroclinic cycles under various (action, memory) pairs.

6 Conclusion

This study contributes to an understanding of a cutting-edge model of learning in games in Sections 3 and 4. In practice, several famous algorithms, i.e., replicator dynamics and gradient ascent, were newly extended to multi-memory games (Algorithms 1 and 2). Then, we proved the correspondence between these algorithms (Theorems 1-3) in general and the uniqueness of Nash equilibrium in two-action zero-sum games (Theorem 4). As a background, multi-agent learning dynamics are generally complicated; thus, many theoretical approaches usually have been taken to grasp such complicated dynamics. In light of this background, our theorems succeeded in capturing the learning dynamics in multi-memory games, which are even more complicated than usual memory-less ones.

This study also experimentally discovered a novel and non-trivial phenomenon that simple learning algorithms such as replicator dynamics and gradient ascent asymptotically reaches a heteroclinic cycle in multi-memory zero-sum games. In other words, the players choose actions in highly skewed proportions throughout learning. Such a phenomenon is specific to multi-memory games: Perhaps this is because the