Algorithm 2 (Discretized MMGA) takes not only its learning rate η but a small value γ in measuring an approximate gradient as inputs. In each time step, each player measures the gradients of its payoff for each variable of its strategy (lines 2-6). Then, the player updates its strategy by the gradients (lines 7-10). Here, note that the strategy update is weighted by the probability $x^{a|s}$ (line 8) in order to correspond to Algorithm 1. Here, each of lines 3-5 and line 8 can be updated in parallel with respect to a and s.

4 Theoretical Analysis

4.1 Continuous-Time Equivalence of Algorithms

The following theorems provide a unified understanding of different algorithms. Theorem 1 and 2 are concerned with continualization of the two discrete algorithms. Surprisingly, Theorem 3 proves the correspondence between these different continualized algorithms by Theorem 1 and 2.

Theorem 1 (Coutinualized MMRD)**.** *Let $\boldsymbol{p}^{a|s}$ be the expected distribution when X chooses a under state s;*

$$p_{i'}^{a|s} := \begin{cases} y^{b|s} & (s_{i'} = abs^{-}) \\ 0 & (\text{otherwise}) \end{cases} . \tag{7}$$

In the limit of $\eta \to 0$, Algorithm 1 is continualized as dynamics

$$\dot{x}^{a|s_i}(\mathbf{x},\mathbf{y}) = p_i^{\mathrm{st}} x^{a|s_i} \left(\pi(\boldsymbol{p}^{a|s_i},\mathbf{x},\mathbf{y}) - \bar{\pi}^{s_i}(\mathbf{x},\mathbf{y}) \right), \tag{8}$$

$$\bar{\pi}^{s_i}(\mathbf{x},\mathbf{y}) = \sum_a x^{a|s_i} \pi(\boldsymbol{p}^{a|s_i},\mathbf{x},\mathbf{y}), \tag{9}$$

for all $a \in \mathcal{A}$ and $s \in \mathcal{S}$. Here, $\bar{\pi}^{s_i}$ is the expected payoff under state s_i.

Theorem 2 (Continualized MMGA)**.** *In the limit of $\gamma \to 0$ and $\eta \to 0$, Algorithm 2 is continualized as dynamics*

$$\dot{x}^{a|s}(\mathbf{x},\mathbf{y}) = x^{a|s} \frac{\partial}{\partial x^{a|s}} u^{\mathrm{st}}(\mathrm{Norm}(\mathbf{x}),\mathbf{y}), \tag{10}$$

for all $a \in \mathcal{A}$ and $s \in \mathcal{S}$.

See Appendix A.1 and A.2 for the proof of Theorems 1 and 2.

Theorem 3 (Equivalence between the algorithms)**.** *The dynamics Eqs.* (8) *and* (10) *are equivalent.*

Proof Sketch. Let $\mathbf{x}'$ be the strategy given by $x^{a|s} \leftarrow x^{a|s} + \gamma$ in $\mathbf{x}$ for $a \in \mathcal{A}$ and $s \in \mathcal{S}$. Then, we consider the changes of the Markov transition matrix $\mathrm{d}\boldsymbol{M} := \boldsymbol{M}(\mathrm{Norm}(\mathbf{x}'),\mathbf{y}) - \boldsymbol{M}(\mathbf{x},\mathbf{y})$ and the stationary distribution $\mathrm{d}\boldsymbol{p}^{\mathrm{st}} := \boldsymbol{p}^{\mathrm{st}}(\mathrm{Norm}(\mathbf{x}'),\mathbf{y}) - \boldsymbol{p}^{\mathrm{st}}(\mathbf{x},\mathbf{y})$. By considering this changes in the stationary condition $\boldsymbol{p}^{\mathrm{st}} = \boldsymbol{M}\boldsymbol{p}^{\mathrm{st}}$, we get $\mathrm{d}\boldsymbol{p}^{\mathrm{st}} = (\boldsymbol{E} - \boldsymbol{M})^{-1}\mathrm{d}\boldsymbol{M}\boldsymbol{p}^{\mathrm{st}}$ in $O(\gamma)$. The right-hand (resp. left-hand) side of this equation corresponds to the continualized MMRD (resp. MMGA). See Appendix A.3 for the full proof. □

For games with a general number of actions, the study [7] has proposed a gradient ascent algorithm in relation to replicator dynamics. In light of this study, Theorem 3 extends the relation to the multi-memory games. This extension is neither simple nor trivial. The relation between replicator dynamics and gradient ascent has been proved by directly calculating $u^{\mathrm{st}} = \boldsymbol{p}^{\mathrm{st}} \cdot \boldsymbol{u}$ [17]. In multi-memory games, however, $u^{\mathrm{st}} = \boldsymbol{p}^{\mathrm{st}} \cdot \boldsymbol{u}$ is too hard to calculate. Thus, as seen in the proof sketch, we proved the relation by considering a slight change in the stationary condition $\boldsymbol{p}^{\mathrm{st}} = \boldsymbol{M}\boldsymbol{p}^{\mathrm{st}}$, technically avoiding such a hard direct calculation.

4.2 Learning Dynamics Near Nash Equilibrium

Below, let us discuss the learning dynamics in multi-memory games, especially divergence from Nash equilibrium in zero-sum payoff matrices. In order to obtain a phenomenological insight into the learning dynamics simply, we assume one-memory two-action zero-sum games in Assumption 1.