STABILITY AND PERFORMANCE ANALYSIS OF DIFFUSION LEARNING FOR TWO-NETWORK COMPETING PROBLEMS

Vladyslav Shashkov, Flávio R. M. Pavan, Haoyuan Cai, and Ali H. Sayed

Adaptive Systems Laboratory, EPFL

ABSTRACT

We address key theoretical questions concerning competing diffusion (CD) algorithms [1], which are designed to solve an important class of network game problems. Specifically, we focus on a competition scenario involving two networks, where each network consists of cooperating agents that are partially connected to the other through a subset of intermediate edges. During each interaction round, the networks simultaneously observe partial information about their opponents and adapt their strategies accordingly. Building on the foundations of the CD algorithm, we present first-, second-, and fourth-order mean stability analyses, as well as a mean-square deviation (MSD) analysis, providing useful performance guarantees for network competition scenarios under some mild assumptions. Computer simulations conducted on quadratic games illustrate our theoretical findings.

Index Terms— Network competition, Nash equilibrium, stability analysis, mean-square deviation (MSD), competing diffusion (CD)

1. INTRODUCTION

Network competition provides great flexibility to model complex game-theoretic interactions among multi-agent systems. This modeling capability lies at the heart of a wide range of real-world applications, such as swarm robotics competition [2], Cournot team game [3], transportation networks [4], and beyond. The rapid growth of existing applications, together with the emergence of new ones, call for the development of efficient algorithms based on solid theoretical analysis.

This work studies an important class of network games involving simultaneous competition between two networks. Prior work [1] introduced the competing diffusion (CD) algorithm to address this problem. The performance of the proposed algorithm was demonstrated mainly through empirical evaluation, leaving theoretical questions open for further investigation. In this work, we develop a theoretical framework to analyze the mean stability and performance at various orders of the CD framework in two-network competition scenarios.

Developing such a framework is nontrivial and is key to understanding the potential of competing algorithms. Our work addresses three primary challenges. First, network game problems capture both cooperative and competitive interactions occurring concurrently, which makes the analysis fundamentally more difficult than existing studies that focus exclusively on either cooperation [5, 6, 7, 8, 9] or competition [10, 11, 12, 13]. Secondly, the MSD analysis has not been pursued in prior game literature and establishing it is not straightforward. Finally, it is rather demanding to determine convergence and stability conditions in scenarios where cross-network connectivity can be extremely weak or sparse.

Some existing related works share similar themes, such as [10, 11, 12, 13, 14, 15, 16, 17, 18]. Our work differs in several aspects, especially in regard to the problem formulation and theoretical guarantees. The work in [16] primarily addresses game problems in single-agent settings, without accounting for the strategic interaction over graphs and across networks. The works [10, 11, 12, 13] did not explicitly consider the possibility that competing agents may implicitly form a team. Although some works consider the networked cooperative game [17], the game dynamics is executed independently within each agent, without involving strategic competition across agents. The most closely related works to ours are those on two-network competition [14, 15, 18]. However, they either focus exclusively on zero-sum formulations or rely on conditions, such as bipartite graphs without isolated nodes for cross-team interactions, which can limit their applicability.

The main contributions of this paper are summarized as follows. We present first-, second-, and fourth-order mean stability analyses of the CD algorithm under a general stochastic game setting. Our theoretical results show that the CD algorithm can converge even under an extremely weak cross-team graph structure, where the two networks are connected by a single pair of nodes. We present an MSD performance analysis in the context of network games—a theoretical contribution that is both of particular interest to the game theory community and has the potential to inspire further analysis in broader competition scenarios.

2. PROBLEM STATEMENT AND CD ALGORITHM

In this section, we present the formulation of the two-network competition problem and provide an overview of the CD algo-

979-8-3503-2411-2/25/$31.00 ©2025 IEEE

rithm [1] employed to address it. Let us consider a collection of K agents split into two teams denoted by the index sets $\mathcal{N}^{(1)} = \{1, \cdots, K_1\}$ and $\mathcal{N}^{(2)} = \{K_1+1, \cdots, K\}$. These teams comprise K_1 and $K_2 = K - K_1$ agents, respectively, and the overall set of agents is denoted by $\mathcal{N} = \mathcal{N}^{(1)} \cup \mathcal{N}^{(2)}$. Each team aims to minimize its own objective defined as follows:

$$\min_{x\in\mathbb{R}^{M_1}} J^{(1)}(x,y), \quad J^{(1)}(x,y) = \sum_{k\in\mathcal{N}^{(1)}} p_k^{(1)} J_k^{(1)}(x,y), \tag{1a}$$

$$\min_{y\in\mathbb{R}^{M_2}} J^{(2)}(x,y), \quad J^{(2)}(x,y) = \sum_{k\in\mathcal{N}^{(2)}} p_k^{(2)} J_k^{(2)}(x,y), \tag{1b}$$

where, for $t \in \{1,2\}$,

$$J_k^{(t)}(x,y) = \mathbb{E}_{\boldsymbol{\xi}_k^{(t)}} Q_k^{(t)}(x,y,\boldsymbol{\xi}_k^{(t)}). \tag{2}$$

Here, $x \in \mathbb{R}^{M_1}$ and $y \in \mathbb{R}^{M_2}$ denote the strategies of networks 1 and 2, respectively, and $Q_k^{(t)}$ denotes the stochastic loss function dependent on the random sample $\boldsymbol{\xi}_k^{(t)}$. The superscript (t) and subscript k are the team and agent indices, respectively. Furthermore, for $t \in \{1,2\}$, it holds that each $p_k^{(t)} > 0$ and $\sum_{k\in\mathcal{N}^{(t)}} p_k^{(t)} = 1$. Note that the two-network zero-sum game [14, 15] is a special case of the problem (1a)–(1b) when $J^{(2)}(x,y) = -J^{(1)}(x,y)$.

2.1. Revisiting the CD Algorithm

The competing diffusion algorithm is briefly revisited here before presenting the associated theoretical analysis in the next section. The algorithm is motivated by the adapt-then-combine (ATC) diffusion learning strategy in distributed optimization [6], incorporating essential elements to handle cross-team information inference. Specifically, beyond the usual ATC step that promotes cooperation within teams, an inference step is introduced to support cross-team information exchange. These core steps, referred to as within-team diffusion and cross-team inference, constitute the main elements of the CD algorithm.

In the phase of within-team diffusion, agents in each team adapt their strategy by running an ATC step, i.e., performing a stochastic gradient step before diffusing the local information to their neighboring agents within the same team. With a slight abuse of notation, let us consider either an agent $k \in \mathcal{N}^{(1)}$ in Team 1 or an agent $k \in \mathcal{N}^{(2)}$ in Team 2. At time instant i, these agents perform the following updates within their local neighborhoods:

$$\boldsymbol{x}_{k,i} = \sum_{\ell\in\mathcal{N}^{(1)}} a_{\ell k}^{(1)} \Big[\boldsymbol{x}_{\ell,i-1} - \mu^{(1)} \widehat{\boldsymbol{\nabla}_x J}_\ell^{(1)}(\boldsymbol{x}_{\ell,i-1}, \boldsymbol{y}'_{\ell,i-1})\Big], \tag{3a}$$

$$\boldsymbol{y}_{k,i} = \sum_{\ell\in\mathcal{N}^{(2)}} a_{\ell k}^{(2)} \Big[\boldsymbol{y}_{\ell,i-1} - \mu^{(2)} \widehat{\boldsymbol{\nabla}_y J}_\ell^{(2)}(\boldsymbol{x}'_{\ell,i-1}, \boldsymbol{y}_{\ell,i-1})\Big], \tag{3b}$$

where $a_{\ell k}^{(1)}$ represents the scaling factor for information flowing from agent ℓ to agent k in Team 1, and $\boldsymbol{y}'_{\ell,i-1}$ denotes the inferred information, at an agent ℓ from Team 1, regarding the strategy of Team 2 at iteration $i-1$. The variables $a_{\ell k}^{(2)}$ and $\boldsymbol{x}'_{\ell,i-1}$ are defined in a similar manner. In streaming data and online learning scenarios, the full gradient for each agent cannot be directly computed and is instead approximated using a stochastic gradient construction. For instance, the stochastic gradient $\widehat{\boldsymbol{\nabla}_x J}_\ell^{(1)}(\boldsymbol{x}_{\ell,i-1}, \boldsymbol{y}'_{\ell,i-1})$ at agent ℓ in Team 1 is computed using a random sample $\boldsymbol{\xi}_{\ell,i}^{(t)}$ and evaluated at the local model $\boldsymbol{x}_{\ell,i-1}$ and inferred opponent's action $\boldsymbol{y}'_{\ell,i-1}$. To enable the competition among networks, both teams have to infer their opponent's strategy. The inference step in the CD algorithm is designed to enable agents to access and respond to adversarial information. In this step, each agent gathers information received from neighboring agents, whether from the same team or the opposing team:

$$\boldsymbol{y}'_{k,i} = \sum_{\ell\in\mathcal{N}^{(2)}} a_{\ell k}^{(21)} \boldsymbol{y}_{\ell,i-1} + \sum_{\ell\in\mathcal{N}^{(1)}} a_{\ell k}^{(11)} \boldsymbol{y}'_{\ell,i-1}, \tag{4a}$$

$$\boldsymbol{x}'_{k,i} = \sum_{\ell\in\mathcal{N}^{(1)}} a_{\ell k}^{(12)} \boldsymbol{x}_{\ell,i-1} + \sum_{\ell\in\mathcal{N}^{(2)}} a_{\ell k}^{(22)} \boldsymbol{x}'_{\ell,i-1}. \tag{4b}$$

Note that the cross-team combination coefficients play a critical role in this process, and we only need a weak condition to ensure cross-team information exchange. The procedure above is summarized in Algorithm 1.

Algorithm 1 Competing diffusion (CD).

---

initialize $i = 0$,
    actions $\{x_{k,-1}, y_{k,-1}, x'_{k,-1}, y'_{k,-1}\}$ for all $k \in \mathcal{N}$
while not done **do**
  for $k \in \mathcal{N}^{(1)}$ **and** $k \in \mathcal{N}^{(2)}$ **in parallel do**

  (within-team diffusion)

$$\boldsymbol{x}_{k,i} = \sum_{\ell\in\mathcal{N}^{(1)}} a_{\ell k}^{(1)} \Big[\boldsymbol{x}_{\ell,i-1} - \mu^{(1)} \widehat{\boldsymbol{\nabla}_x J}_\ell^{(1)}(\boldsymbol{x}_{\ell,i-1}, \boldsymbol{y}'_{\ell,i-1})\Big]$$

$$\boldsymbol{y}_{k,i} = \sum_{\ell\in\mathcal{N}^{(2)}} a_{\ell k}^{(2)} \Big[\boldsymbol{y}_{\ell,i-1} - \mu^{(2)} \widehat{\boldsymbol{\nabla}_y J}_\ell^{(2)}(\boldsymbol{x}'_{\ell,i-1}, \boldsymbol{y}_{\ell,i-1})\Big]$$

  (cross-team inference)

$$\boldsymbol{y}'_{k,i} = \sum_{\ell\in\mathcal{N}^{(2)}} a_{\ell k}^{(21)} \boldsymbol{y}_{\ell,i-1} + \sum_{\ell\in\mathcal{N}^{(1)}} a_{\ell k}^{(11)} \boldsymbol{y}'_{\ell,i-1}$$

$$\boldsymbol{x}'_{k,i} = \sum_{\ell\in\mathcal{N}^{(1)}} a_{\ell k}^{(12)} \boldsymbol{x}_{\ell,i-1} + \sum_{\ell\in\mathcal{N}^{(2)}} a_{\ell k}^{(22)} \boldsymbol{x}'_{\ell,i-1}$$

  end for
  $i \leftarrow i + 1$
end while

---

3. THEORETICAL ANALYSIS

This section presents the theoretical results for the CD algorithm. The main results are split into two theorems which include stability results for the weight error of first, second, and fourth orders, and the MSD performance results, respectively. For convenience of analysis, we introduce the composed two-network global gradient mapping

$$F(z) = \mathrm{col}\{\nabla_x J^{(1)}(x,y), \nabla_y J^{(2)}(x,y)\}, \tag{5}$$

where $z = \mathrm{col}\{x, y\} \in \mathbb{R}^{M_1+M_2}$ and $\mathrm{col}\{\cdot\}$ denotes the column stacking of vectors. In addition, the short notation $J_k^{(t)}(z) = J_k^{(t)}(x,y)$ for the risk value is used when necessary.

Before the theoretical results are established, some standard assumptions are introduced in the following.

Assumption 1 (Within- and cross-team combination matrices). Let $A^{(1)} \in \mathbb{R}^{K_1\times K_1}$ and $A^{(2)} \in \mathbb{R}^{K_2\times K_2}$ denote the within-team combination matrices, and let $A_{\mathrm{blk}}^{(21)} \in \mathbb{R}^{K\times K_1}$ and $A_{\mathrm{blk}}^{(12)} \in \mathbb{R}^{K\times K_2}$ denote the cross-team combination matrices. We assume that the entries of the combination matrices $A^{(1)}$ and $A^{(2)}$ are nonnegative and that these matrices are:

- **(i)** *Left-stochastic*: $\mathbb{1}_{K_1}^\mathsf{T} A^{(1)} = \mathbb{1}_{K_1}^\mathsf{T}$ and $\mathbb{1}_{K_2}^\mathsf{T} A^{(2)} = \mathbb{1}_{K_2}^\mathsf{T}$.
- **(ii)** *Primitive*: There exists a positive integer n such that all entries of $(A^{(1)})^n$ and $(A^{(2)})^n$ are strictly positive.

On the other hand, the entries of the cross-team combination matrices are nonnegative and these matrices are:

- **(iii)** *Left-stochastic*: $\mathbb{1}_{K}^\mathsf{T} A_{\mathrm{blk}}^{(21)} = \mathbb{1}_{K_1}^\mathsf{T}$ and $\mathbb{1}_{K}^\mathsf{T} A_{\mathrm{blk}}^{(12)} = \mathbb{1}_{K_2}^\mathsf{T}$.
- **(iv)** *Semi-weakly positive*: Consider the partitioned form of the cross-team combination matrices as follows:

$$A_{\mathrm{blk}}^{(21)} = \begin{bmatrix} A^{(21)} \\ A^{(11)} \end{bmatrix} \quad \text{and} \quad A_{\mathrm{blk}}^{(12)} = \begin{bmatrix} A^{(12)} \\ A^{(22)} \end{bmatrix} \tag{6}$$

where $A^{(tt')} \in \mathbb{R}^{K_t\times K_{t'}}$ and $A^{(tt)} \in \mathbb{R}^{K_t\times K_t}$. It is assumed that the upper block matrices $A^{(tt')}$ have at least one positive entry. Furthermore, $A^{(tt)}$ are primitive. Consequently, there exists a path from every agent in Team 2 to every agent in Team 1, and vice versa. □

As the presentation will reveal, the conditions above on the combination matrices enable consensus among teammates and allow information inference from opponents with minimal connectivity. Furthermore, the local risk functions are assumed to satisfy the conditions that follow.

Assumption 2 (Step size condition). Assume that the step sizes for the algorithm are chosen sufficiently small such that

$$4\frac{\mu^{(2)}}{\mu^{(1)}}\nu_{\mathrm{eff}}^{(1)}\nu_{\mathrm{eff}}^{(2)} - \left(\delta_{\mathrm{eff}}^{(21)} + \frac{\mu^{(2)}}{\mu^{(1)}}\delta_{\mathrm{eff}}^{(12)}\right)^2 \geq \epsilon, \tag{7}$$

where $\epsilon > 0$ is an arbitrarily small parameter for which

$$\sum_{k\in\mathcal{N}^{(1)}} p_k^{(1)} \nabla_x^2 J_k^{(1)}(x_k, y_k) \geq \epsilon, \quad \forall\ \{x_k, y_k\}_{k\in\mathcal{N}^{(1)}}, \tag{8a}$$

$$\sum_{k\in\mathcal{N}^{(2)}} p_k^{(2)} \nabla_y^2 J_k^{(2)}(x_k, y_k) \geq \epsilon, \quad \forall\ \{x_k, y_k\}_{k\in\mathcal{N}^{(2)}}, \tag{8b}$$

and

$$\delta_{\mathrm{eff}}^{(t't)} = \sup_{\{x_k,y_k\}_{k\in\mathcal{N}^{(t)}}} \Big\| \sum_{k\in\mathcal{N}^{(t)}} p_k^{(t)} \nabla_{xy}^2 J_k^{(t)}(x_k, y_k) \Big\|, \tag{9a}$$

$$\nu_{\mathrm{eff}}^{(1)} = \inf_{\{x_k,y_k\}_{k\in\mathcal{N}^{(1)}}} \Big\| \sum_{k\in\mathcal{N}^{(1)}} p_k^{(1)} \nabla_x^2 J_k^{(1)}(x_k, y_k) \Big\|, \tag{9b}$$

$$\nu_{\mathrm{eff}}^{(2)} = \inf_{\{x_k,y_k\}_{k\in\mathcal{N}^{(2)}}} \Big\| \sum_{k\in\mathcal{N}^{(2)}} p_k^{(2)} \nabla_y^2 J_k^{(2)}(x_k, y_k) \Big\|. \tag{9c}$$

□

Conditions (8a)–(8b) and (9a)–(9c) imply strong monotonicity on the gradient mapping $F(z)$. In the context of the game problem (1a)–(1b), this condition implies the existence and uniqueness of the Nash equilibrium point $z^\star = \mathrm{col}\{x^\star, y^\star\} \in \mathbb{R}^M$ characterized by the inequalities [19]:

$$J^{(1)}(x^\star, y^\star) \leq J^{(1)}(x, y^\star), \quad \forall\ x \in \mathbb{R}^{M_1}, \tag{10a}$$

$$J^{(2)}(x^\star, y^\star) \leq J^{(2)}(x^\star, y), \quad \forall\ y \in \mathbb{R}^{M_2}. \tag{10b}$$

With a well-defined Nash equilibrium, the stability analysis of the error moment and the MSD analysis can be conducted by appropriately defining the weight error.

Assumption 3 (Local risk functions). For all $k \in \mathcal{N}^{(t)}$, we assume the local risk functions $J_k^{(t)}(\cdot,\cdot)$ satisfy:

- **(i)** *Twice continuous differentiability* with respect to both arguments, meaning that all first-order and second-order partial derivatives exist and are continuous.
- **(ii)** *Smoothness*, i.e., for all vectors $z_1 = \mathrm{col}\{x_1, y_1\},\ z_2 = \mathrm{col}\{x_2, y_2\} \in \mathbb{R}^{M_1+M_2}$, the inequalities

$$\|\nabla_x J_k^{(1)}(z_1) - \nabla_x J_k^{(1)}(z_2)\| \leq \delta_k^{(1)}(\|z_1 - z_2\|), \tag{11a}$$

$$\|\nabla_y J_k^{(2)}(z_1) - \nabla_y J_k^{(2)}(z_2)\| \leq \delta_k^{(2)}(\|z_1 - z_2\|) \tag{11b}$$

hold for some $\delta_k^{(1)}, \delta_k^{(2)} > 0$.

- **(iii)** *Hessian is locally Lipschitz continuous*, i.e., the inequalities

$$\|\nabla_x^2 J_k^{(1)}(z^\star + \Delta z) - \nabla_x^2 J_k^{(1)}(z^\star)\| \leq \kappa_{\mathrm{d}}\|\Delta z\|, \tag{12a}$$

$$\|\nabla_y^2 J_k^{(2)}(z^\star + \Delta z) - \nabla_y^2 J_k^{(2)}(z^\star)\| \leq \kappa_{\mathrm{d}}\|\Delta z\|, \tag{12b}$$

$$\|\nabla_{xy}^2 J_k^{(t)}(z^\star + \Delta z) - \nabla_{xy}^2 J_k^{(t)}(z^\star)\| \leq \kappa_{\mathrm{d}}\|\Delta z\| \tag{12c}$$

hold for a small $\|\Delta z\| \leq \epsilon$ and some $\kappa_{\mathrm{d}} > 0$. □

The first two conditions are standard assumptions in the diffusion learning context [6], implying

$$\|\nabla_x^2 J_k^{(1)}\| \leq \delta_k^{(1)}, \qquad \|\nabla_{xy}^2 J_k^{(1)}\| \leq \delta_k^{(1)}, \tag{13a}$$

$$\|\nabla_y^2 J_k^{(2)}\| \leq \delta_k^{(2)}, \qquad \|\nabla_{xy}^2 J_k^{(2)}\| \leq \delta_k^{(2)}. \tag{13b}$$

The third condition, which is mild since it is merely imposed around the neighborhood of $z^\star$, is usually employed to establish the first-order stability of stochastic algorithms [6].

Assumption 4 (**Gradient noise processes**). Let

$$\boldsymbol{\mathcal{F}}_{i-1} = \bigcup_{j\leq i-1} \{\boldsymbol{x}_{k,j}, \boldsymbol{y}'_{k,j}\}_{k\in\mathcal{N}^{(1)}} \cup \{\boldsymbol{x}'_{k,j}, \boldsymbol{y}_{k,j}\}_{k\in\mathcal{N}^{(2)}} \tag{14}$$

denote the filtration generated by random processes in both networks. For all $k \in \mathcal{N}^{(t)}$ and $t \in \{1, 2\}$, we assume the gradient noises defined by

$$\boldsymbol{s}_{k,i}^{(1)} = \widehat{\boldsymbol{\nabla}_x J}_k^{(1)}(\boldsymbol{x}_{k,i-1}, \boldsymbol{y}'_{k,i-1}) - \nabla_x J_k^{(1)}(\boldsymbol{x}_{k,i-1}, \boldsymbol{y}'_{k,i-1}), \tag{15a}$$

$$\boldsymbol{s}_{k,i}^{(2)} = \widehat{\boldsymbol{\nabla}_y J}_k^{(2)}(\boldsymbol{x}'_{k,i-1}, \boldsymbol{y}_{k,i-1}) - \nabla_y J_k^{(2)}(\boldsymbol{x}'_{k,i-1}, \boldsymbol{y}_{k,i-1}) \tag{15b}$$

satisfy the following conditions:

$$\mathbb{E}(\boldsymbol{s}_{k,i}^{(t)} \,|\, \boldsymbol{\mathcal{F}}_{i-1}) = 0, \tag{16a}$$

$$\mathbb{E}(\boldsymbol{s}_{k,i}^{(t)} \boldsymbol{s}_{\ell,i}^{(t)\mathsf{T}} \,|\, \boldsymbol{\mathcal{F}}_{i-1}) = 0, \quad \forall\, k \neq \ell, \tag{16b}$$

$$\mathbb{E}(\|\boldsymbol{s}_{k,i}^{(t)}\|^4 \,|\, \boldsymbol{\mathcal{F}}_{i-1}) \leq (\bar{\beta}_k^{(t)})^4 \|\boldsymbol{z}_{k,i-1}\|^4 + (\bar{\sigma}_k^{(t)})^4 \tag{16c}$$

for some $\bar{\beta}_k^{(t)}, \bar{\sigma}_k^{(t)} \geq 0$. □

The assumption that follows will be employed in the MSD performance analysis.

Assumption 5 (**Noise covariance matrix**). Let $s_{k,i}^{(t)}(z)$ be the gradient noise evaluated at $z = \text{col}\{x, y\}$. We assume the gradient noise processes of all agents satisfy

$$\Big| \lim_{i\to+\infty} \mathbb{E}(\boldsymbol{s}_{k,i}^{(t)}(z^\star + \Delta z)\, \boldsymbol{s}_{k,i}^{(t)\mathsf{T}}(z^\star + \Delta z) \,|\, \boldsymbol{\mathcal{F}}_{i-1}) - \lim_{i\to+\infty} \mathbb{E}(\boldsymbol{s}_{k,i}^{(t)}(z^\star)\, \boldsymbol{s}_{k,i}^{(t)\mathsf{T}}(z^\star) \,|\, \boldsymbol{\mathcal{F}}_{i-1}) \Big| \leq L\|\Delta z\|^\gamma \tag{17}$$

for some $\gamma \in (0, 4]$, a positive constant $L > 0$, and small perturbations $\|\Delta z\| \leq \epsilon$. □

Our stability and performance analysis relies on an error recursion characterized by the deviation between the current iterates and the solution point at two successive iterations. In order to derive it, we start from the networked recursion and subtract the solution point from both sides. In the following, we state the recursion and omit its lengthy derivation:

$$\widetilde{\boldsymbol{z}}_i = \boldsymbol{\mathcal{B}}_{i-1}\widetilde{\boldsymbol{z}}_{i-1} + \mathcal{A}^\mathsf{T}\mathcal{M}\boldsymbol{s}_i + \mathcal{A}^\mathsf{T}\mathcal{M}b, \tag{18}$$

where

$$\widetilde{\boldsymbol{z}}_i = \text{col}\{x^\star \otimes \mathbb{1}_K, y^\star \otimes \mathbb{1}_K\} - \boldsymbol{z}_i, \tag{19a}$$

$$\boldsymbol{z}_i = \text{col}\{\{\boldsymbol{x}_{k,i}\}_{k\in\mathcal{N}^{(1)}}, \{\boldsymbol{x}'_{k,i}\}_{k\in\mathcal{N}^{(2)}}, \{\boldsymbol{y}_{k,i}\}_{k\in\mathcal{N}^{(2)}}, \{\boldsymbol{y}'_{k,i}\}_{k\in\mathcal{N}^{(1)}}\}, \tag{19b}$$

$$\boldsymbol{\mathcal{B}}_{i-1} = \Theta^\mathsf{T} - \mathcal{A}^\mathsf{T}\mathcal{M}\boldsymbol{\mathcal{H}}_{i-1}, \tag{19c}$$

$$\Theta = \text{blkdiag}\left\{ \begin{bmatrix} A^{(1)} \otimes I_{M_1} & A^{(12)} \otimes I_{M_1} \\ 0_{M_1K_2\times M_1K_1} & A^{(22)} \otimes I_{M_1} \end{bmatrix}, \begin{bmatrix} A^{(2)} \otimes I_{M_2} & A^{(21)} \otimes I_{M_2} \\ 0_{M_2K_1\times M_2K_2} & A^{(11)} \otimes I_{M_2} \end{bmatrix} \right\}, \tag{19d}$$

$$\mathcal{A} = \text{blkdiag}\left\{ \begin{bmatrix} A^{(1)} \otimes I_{M_1} & 0_{M_1K_1\times M_1K_2} \\ 0_{M_1K_2\times M_1K_1} & 0_{M_1K_2\times M_1K_2} \end{bmatrix}, \begin{bmatrix} A^{(2)} \otimes I_{M_2} & 0_{M_2K_2\times M_2K_1} \\ 0_{M_2K_1\times M_2K_2} & 0_{M_2K_1\times M_2K_1} \end{bmatrix} \right\}, \tag{19e}$$

$$\mathcal{M} = \text{diag}\{\mu^{(1)}\mathbb{1}_{M_1K}, \mu^{(2)}\mathbb{1}_{M_2K}\}, \tag{19f}$$

$$\boldsymbol{s}_i = \text{col}\{\{\boldsymbol{s}_{k,i}^{(1)}\}_{k\in\mathcal{N}^{(1)}}, 0_{M_1K_2}, \{\boldsymbol{s}_{k,i}^{(2)}\}_{k\in\mathcal{N}^{(2)}}, 0_{M_2K_1}\}, \tag{19g}$$

$$b = \text{col}\{\{\nabla_x J_k^{(1)}(z^\star)\}_{k\in\mathcal{N}^{(1)}}, 0_{M_1K_2}, \{\nabla_y J_k^{(2)}(z^\star)\}_{k\in\mathcal{N}^{(2)}}, 0_{M_2K_1}\}, \tag{19h}$$

and

$$\boldsymbol{\mathcal{H}}_i = \begin{bmatrix} \boldsymbol{\mathcal{H}}_i^{(1)} & 0 & 0 & \boldsymbol{\mathcal{H}}_i^{(21)} \\ 0 & 0 & 0 & 0 \\ 0 & \boldsymbol{\mathcal{H}}_i^{(12)} & \boldsymbol{\mathcal{H}}_i^{(2)} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \tag{20a}$$

$$\boldsymbol{\mathcal{H}}_i^{(t)} = \text{blkdiag}\{\boldsymbol{H}_{k,i}^{(t)} : k \in \mathcal{N}^{(t)}\}, \tag{20b}$$

$$\boldsymbol{\mathcal{H}}_i^{(t't)} = \text{blkdiag}\{\boldsymbol{H}_{k,i}^{(t't)} : k \in \mathcal{N}^{(t)}\}. \tag{20c}$$

The matrices $\boldsymbol{H}_{k,i}^{(t)}$ are defined as definite integrals of Hessians according to [6, Eq. (8.171)], but with the integrand replaced by $\nabla_x^2 J_k^{(1)}$ if $t = 1$ and by $\nabla_y^2 J_k^{(2)}$ if $t = 2$. Likewise, the matrices $\boldsymbol{H}_{k,i}^{(t't)}$ follow the same definition, but with the integrand replaced by $\nabla_{yx}^2 J_k^{(1)}$ if $(t, t') = (1, 2)$ and by $\nabla_{xy}^2 J_k^{(2)}$ if $(t, t') = (2, 1)$.

Theorem 1 (**Stability results**). *Let Assumptions 1–4 hold. The weight error $\widetilde{\boldsymbol{z}}_i$ produced by running CD is mean-stable in the first-, second-, and fourth-order moments under sufficiently small step-sizes, namely,*

$$\limsup_{i\to+\infty} \mathbb{E}\,\|\widetilde{\boldsymbol{z}}_i\|^2 = O(\mu_{\max}), \tag{21a}$$

$$\limsup_{i\to+\infty} \mathbb{E}\,\|\widetilde{\boldsymbol{z}}_i\|^4 = O(\mu_{\max}^2), \tag{21b}$$

$$\limsup_{i\to+\infty} \|\mathbb{E}\,\widetilde{\boldsymbol{z}}_i\| = O(\mu_{\max}), \tag{21c}$$

where $\mu_{\max} = \max\{\mu^{(1)}, \mu^{(2)}\}$.

Proof. The derivation needs to extend arguments similar to those used in [6]. For brevity, we summarize the main steps. To obtain the first two results, one needs to consider the Jordan canonical form of the matrix $\mathit{\Theta} = \mathcal{V}\mathcal{J}\mathcal{V}^{-1}$ and transform recursion (18) into a recursion for $\mathcal{V}^{\mathsf{T}}\widetilde{\boldsymbol{z}}_i$. Moreover, the analysis needs to handle two parts, one related to the noisy term and the other associated with $\mathcal{V}^{\mathsf{T}}\widetilde{\boldsymbol{z}}_{i-1}$. Using Assumption 4 and the fact that the transition matrix multiplying the previous iterate $\mathcal{V}^{\mathsf{T}}\widetilde{\boldsymbol{z}}_{i-1}$ decays out over iterations, the moment $\mathbb{E}\,\|\mathcal{V}^{\mathsf{T}}\widetilde{\boldsymbol{z}}_i\|^{\alpha}$ can be upper bounded. For the analysis of the first-order error moment, we need to verify the stability of $\boldsymbol{\mathcal{B}}_{i-1}$ when it is evaluated at $z^\star$ and the subsequent analysis can be closed by invoking Assumption 2. □

Theorem 2 (Performance analysis). *Under Assumptions 1–5, it holds that*

$$\frac{1}{4K}\limsup_{i\to+\infty}\mathbb{E}\,\|\widetilde{\boldsymbol{z}}_i\|^2 = \frac{1}{4K}\,\mathrm{Tr}(\mathcal{X}) + o(\mu_{\max}), \tag{22a}$$

$$\limsup_{i\to+\infty}\mathbb{E}\,\|\boldsymbol{x}_{k,i} - x^\star\|^2 = \mathrm{Tr}(\mathcal{J}_k^{(1)}\mathcal{X}) + o(\mu_{\max}), \tag{22b}$$

$$\limsup_{i\to+\infty}\mathbb{E}\,\|\boldsymbol{y}_{k,i} - y^\star\|^2 = \mathrm{Tr}(\mathcal{J}_k^{(2)}\mathcal{X}) + o(\mu_{\max}) \tag{22c}$$

where

$$\mathcal{X} = \left(\sum_{n=0}^{+\infty}\mathcal{B}^{n\mathsf{T}}\mathcal{B}^n\right)\mathcal{A}^{\mathsf{T}}\mathcal{M}\mathcal{S}\mathcal{M}\mathcal{A}, \tag{23a}$$

$$\mathcal{B} = \mathit{\Theta}^{\mathsf{T}} - \mathcal{A}^{\mathsf{T}}\mathcal{M}\mathcal{H}, \tag{23b}$$

$$\mathcal{S} = \mathrm{blkdiag}\{S^{(1)}, 0_{M_1K_2\times M_1K_2}, S^{(2)}, 0_{M_2K_1\times M_2K_1}\}, \tag{23c}$$

$$\mathcal{S}^{(t)} = \mathrm{blkdiag}\left\{\lim_{i\to\infty}\mathbb{E}(\boldsymbol{s}_{k,i}^{(t)}(z^\star)\,\boldsymbol{s}_{k,i}^{(t)\mathsf{T}}(z^\star)\,|\,\boldsymbol{\mathcal{F}}_{i-1}) : k\in\mathcal{N}^{(t)}\right\}, \tag{23d}$$

$$\mathcal{J}_k^{(1)} = \mathrm{diag}\{0_{M_1k-M_1}, \mathbb{1}_{M_1}, 0_{M_1K-M_1k}, 0_{M_2K}\}, \tag{23e}$$

$$\mathcal{J}_k^{(2)} = \mathrm{diag}\{0_{M_1K}, 0_{M_2k-M_2}, \mathbb{1}_{M_2}, 0_{M_2K-M_2k}\}. \tag{23f}$$

In (23b)*, $\mathcal{H}$ denotes the deterministic version of $\boldsymbol{\mathcal{H}}_{i-1}$ and is defined according to* (20a)*, but evaluated at the Nash equilibrium. Finally, the convergence rate of $\mathbb{E}\,\|\widetilde{\boldsymbol{z}}_i\|^2$ in* (22a) *is* $\alpha = 1 - \Omega(\mu_{\max})$.

Proof. The proof for performance results relies on a long-term dynamics version of (18), which is given by

$$\widetilde{z}_i' = \mathcal{B}\widetilde{z}_{i-1}' + \mathcal{A}^{\mathsf{T}}\mathcal{M}\boldsymbol{s}_i + \mathcal{A}^{\mathsf{T}}\mathcal{M}b. \tag{24}$$

For a long sequence of iterations, one can show that the long-term dynamics of the error recursion remains close to the original model within a very small neighborhood, i.e.,

$$\limsup_{i\to+\infty}\mathbb{E}\,\|\widetilde{\boldsymbol{z}}_i - \widetilde{z}_i'\|^2 = O\left(\mu_{\max}^2\right). \tag{25}$$

The analysis focuses, therefore, on a more tractable $\widetilde{z}_i'$. Due to space constraints, we omit the details of the proof. □

4. SIMULATION RESULTS

In order to validate our theoretical findings, we present simulation results for a quadratic game. Such competition problems are particularly relevant in the context of network games in economics [20]. In the setting considered next, the local risk function at agent k is given by

$$J_k^{(t)}(z) = \frac{1}{2}\,\mathbb{E}\,\|\boldsymbol{u}_{k,i}^{\mathsf{T}}z - \boldsymbol{d}(k,i)\|^2, \tag{26}$$

where $\{\boldsymbol{u}_{k,i}, \boldsymbol{d}(k,i)\}$ are the observations at agent k. In this experiment, these observations are generated according to the random processes $\boldsymbol{u}_{k,i}\sim N(0, R_{k,u})$, $\boldsymbol{v}(k,i)\sim N(0,\sigma_{k,v}^2)$ and $\boldsymbol{d}(k,i) = \boldsymbol{u}_{k,i}z_k^\star + \boldsymbol{v}(k,i)$, where $z_k^\star\in\mathbb{R}^M$ is the solution vector, $R_{k,u}\in\mathbb{R}^{M\times M}$, $\sigma_{k,v}\in\mathbb{R}_{\geq 0}$ are the problem parameters, and $\boldsymbol{v}(k,i)$ is the independent white noise. For the settings above, we can verify that there exists a unique Nash equilibrium. We additionally consider $M_1 = 5$, $M_2 = 10$, $K_1 = 2$, and $K_2 = 4$. The simulation results of first- and fourth-order error moments and the MSD performance are plotted in Fig. 1. We observe that CD is mean-stable in first- and fourth-order moments. Moreover, smaller error moments

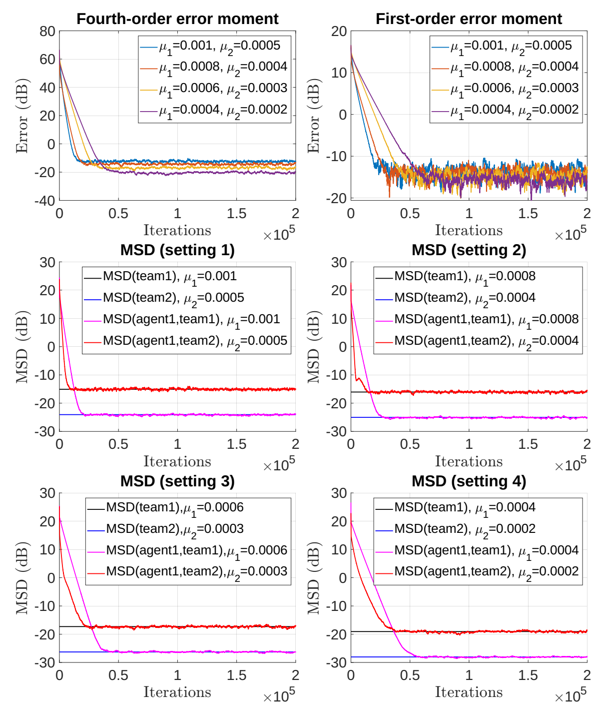

Fig. 1. Simulation results for the first- and fourth-order error moments and the MSD performance. The first plot depicts the fourth-order error curves over several step-sizes, the second plot depicts the first-order error curves over several step-sizes, the third to sixth plot depict the MSD error curves of the first agent of each team compared to the theoretical performance levels. Each MSD plot depicts a separate step-size configuration.

can be obtained via choosing a smaller step size. More importantly, the MSD results match with the theoretical predictions at both the agent and team levels.

5. CONCLUSION

In this paper, we established first-, second-, and fourth-order mean stability results, as well as MSD results, for the competing diffusion (CD) algorithm. Our theoretical results allow for the better understanding of the steady-state behavior of competing networks. In particular, we established a performance analysis of the CD algorithm—an important step that has not yet been conducted in prior literature. Simulations of a quadratic game were presented to validate the theoretical claims.

Acknowledgment. The authors thank Dr. S. Vlaski for useful discussions on the topic of this manuscript.

6. REFERENCES

[1] S. Vlaski and A. H. Sayed, "Competing adaptive networks," in *Proc. IEEE Statistical Signal Processing Workshop (SSP)*, July 2021, pp. 71–75.

[2] Z. He, Y. Sun, and Z. Feng, "Research on resource allocation of autonomous swarm robots based on game theory," *Electronics*, vol. 12, no. 20, 2023, (article number: 4370).

[3] S. S. Askar, "Complex dynamic properties of cournot duopoly games with convex and log-concave demand function," *Operations Research Letters*, pp. 85–90, 2014.

[4] R. Z. Farahani, E. Miandoabchi, W. Y. Szeto, and H. Rashidi, "A review of urban transportation network design problems," *European Journal on Operational Research*, pp. 281–302, 2013.

[5] J. Chen and A. H. Sayed, "Diffusion adaptation strategies for distributed optimization and learning over networks," *IEEE Transactions on Signal Processing*, vol. 60, no. 8, pp. 4289–4305, 2012.

[6] A. H. Sayed, "Adaptation, learning, and optimization over networks," *Foundations and Trends in Machine Learning*, vol. 7, no. 4–5, pp. 311–801, 2014.

[7] A. Nedić and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Transactions on Automatic Control*, vol. 54, no. 1, pp. 48–61, 2009.

[8] D. Jakovetić, J. Xavier, and J. M. F. Moura, "Fast distributed gradient methods," *IEEE Transactions on Automatic Control*, vol. 59, no. 5, pp. 1131–1146, 2014.

[9] R. Xin, A. K. Sahu, U. A. Khan, and S. Kar, "Distributed stochastic optimization with gradient tracking over strongly-connected networks," in *Proc. IEEE Conference on Decision and Control (CDC)*, Dec. 2019, pp. 8353–8358.

[10] M. Ye, L. Yin, G. Wen, and Y. Zheng, "On distributed Nash equilibrium computation: Hybrid games and a novel consensus-tracking perspective," *IEEE Transactions on Cybernetics*, vol. 51, no. 10, pp. 5021–5031, 2020.

[11] K. Lu, "Online distributed algorithms for online noncooperative games with stochastic cost functions: High probability bound of regrets," *IEEE Transactions on Automatic Control*, vol. 69, pp. 8860–8867, 2024.

[12] T. Tatarenko and A. Nedić, "Geometric convergence of distributed gradient play in games with unconstrained action sets," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 3367–3372, Apr. 2020.

[13] F. Schaefer and A. Anandkumar, "Competitive gradient descent," in *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, Dec. 2019, vol. 32, pp. 1–11.

[14] Y. Lou, Y. Hong, L. Xie, G. Shi, and K. H. Johansson, "Nash equilibrium computation in subnetwork zero-sum games with switching communications," *IEEE Transactions on Automatic Control*, vol. 61, no. 10, pp. 2920–2935, 2015.

[15] S. Huang, J. Lei, Y. Hong, U. V. Shanbhag, and J. Chen, "No-regret distributed learning in subnetwork zero-sum games," *IEEE Transactions on Automatic Control*, vol. 69, no. 10, pp. 6620–6635, 2024.

[16] T. Lin, C. Jin, and M. I. Jordan, "On gradient descent ascent for nonconvex-concave minimax problems," in *Proc. International Conference on Machine Learning (PMLR)*, July 2020, vol. 119, pp. 6083–6093.

[17] H. Cai, S. A. Alghunaim, and A. H. Sayed, "Diffusion stochastic optimization for min-max problems," *IEEE Transactions on Signal Processing*, vol. 73, pp. 259–274, 2025.

[18] Y. Zhao, H. Cai, and A. H. Sayed, "Diffusion learning over adaptive competing networks," in *Proc. ICASSP, Hyderabad, India*, 2025, pp. 1–5.

[19] F. Facchinei and J.-S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Springer, 2003.

[20] C. M. Leister, "Information acquisition and welfare in network games," *Games and Economic Behavior*, vol. 122, pp. 453–475, 2020.