DOT and DOP: Linearly Convergent Algorithms for Finding Fixed Points of Multiagent Operators

Xiuxian Li, Senior Member, IEEE, Min Meng, and Lihua Xie, Fellow, IEEE

Abstract—This article investigates the distributed fixed-point finding problem for a global operator over a directed and unbalanced multiagent network, where the global operator is quasi-nonexpansive and only partially accessible to each individual agent. Two cases are addressed, i.e., the global operator is sum separable and block separable. For this first case, the global operator is the sum of local operators, which are assumed to be Lipschitz, and each local operator is privately known to each individual agent. To deal with this scenario, a distributed (or decentralized) algorithm, called distributed quasi-averaged operator tracking algorithm (DOT), is proposed and analyzed, and it is shown that the algorithm can converge to a fixed point of the global operator at a linear rate under a linear regularity condition, which is strictly weaker than the strong convexity assumption on cost functions in existing convex optimization literature. For the second scenario, the global operator is composed of a group of local block mappings which are Lipschitz and can be accessed only by each individual agent. In this setup, a distributed algorithm, called distributed quasi-averaged operator playing algorithm (DOP), is developed and shown to be linearly convergent to a fixed point of the global operator under the linear regularity condition. The above studied problems provide a unified framework for many interesting problems. As examples, the proposed DOT and DOP are exploited to cope with distributed optimization and multiplayer games under partial-decision information. Finally, numerical examples are presented to corroborate the theoretical results.

Index Terms—Distributed algorithms, distributed optimization, fixed point, game, linear convergence, linear regularity, multiagent networks, real Hilbert spaces.

I. INTRODUCTION

FIXED point theory in real Hilbert spaces is known as a powerful tool in a variety of domains such as optimization, engineering, economics, game theory, and nonlinear numerical analysis [1], [2]. Generally speaking, the main goal is to devise algorithms for computing a fixed point of an operator.

Up to now, plenty of research has extensively addressed centralized algorithms for finding a fixed point of nonexpansive or quasi-nonexpansive operators in the literature [3], [4], [5], [6], where a central/global coordinator or computing unit is able to access all information of the studied problem. It is known that the typical Picard iteration in general does not converge for nonexpansive operators (e.g., a simple example is the operator $-Id$ with nonzero initial points, where $Id$ is the identity operator), although it usually performs well for contractive operators. For nonexpansive operators, one prominent algorithm is the so-called Krasnosel’ ski˘ı-Mann (KM) iteration [7], [8], and it is shown to converge weakly to a fixed point of a nonexpansive operator in real Hilbert spaces under mild conditions [9].

In recent few decades, distributed (or decentralized) algorithms have been an active topic in a wide range of domains, including fixed point theory, computer science, game theory, and control theory, and so on, mostly inspired by the fact that distributed algorithms, in contrast with centralized ones, possess a host of fascinating advantages, such as low cost, robustness to failures or antagonistic attacks, privacy preservation, low computational complexity, and so on. Distributed algorithms do not assume global/central coordinators or computing units and instead a finite group of agents (e.g., computing units and robots, and so on), who may be spatially separated, aim to solve a global problem in a collaborative manner by local information exchanges. In this setting, local information exchanges are often depicted by a simple graph, connoting that every agent can interact with only a subset of agents, instead of all agents. Along this line, distributed algorithms have thus far been investigated extensively under both fixed and time-varying communication graphs in distributed optimization [10], [11], [12], game theory [13], and multiagent systems/networks [14], [15], to just name a few.

In more recent years, distributed algorithms have received a growing attention in the fixed-point finding problem [16], [17], [18], [19], [20], [21]. For instance, a synchronous distributed algorithm was proposed in [16] for computing a common fixed point of a collection of paracontractive operators, and for the same problem, an asynchronous distributed
algorithm was developed in [17]. Meanwhile, different from paracontractive operators, another type of operators (i.e., strongly quasi-nonexpansive operators) was addressed for the common fixed-point seeking problem in [18] by designing a distributed algorithm in the presence of time-varying delays under the assumption that the communication graph is repeatedly jointly strongly connected. It should be noted that many interesting problems can boil down to the common fixed-point finding problem, such as convex feasibility problems [22], [23] and the problem of solving linear algebraic equations in a distributed fashion [24], [25], [26], and so forth. For example, the linear algebraic equation solving problem was formulated such that the distribution of random communication graphs is not required, which include asynchronous updates and/or unreliable interconnection protocols. Notice that all the aforementioned works are in the Euclidean space. Observe that the real Hilbert space is more general, which can be both finite and infinite dimensional, including the Euclidean space as a special case, and the corresponding distributed algorithms have been considered recently [19], [20], [21]. However, the infinite dimensional spaces are usually more challenging than the finite ones, e.g., weak convergence and strong convergence are not equivalent in infinite dimensional Hilbert spaces, although equivalent in finite ones (cf. [1], [2]). Along this line, the authors in [19] investigated distributed optimization under random and directed interconnection graphs, where a distributed algorithm was proposed and shown to be convergent in both almost surely and mean square sense along with the introduction of a novel convex minimization problem over the fixed-value point set of a nonexpansive random operator. In addition, the authors in [20] took into account the common fixed-point finding problem for a finite collection of nonexpansive operators, where two distributed algorithms were proposed with a full coordinate updating and a random block-coordinate updating, respectively, and compared with [16], [17], and [18], the contributions of [20] lie in the study of real Hilbert spaces, the consideration of operator errors, and the establishment of a sublinear convergence speed. Furthermore, a more general scenario, where no common fixed points are assumed for all local operators, was investigated in [21], where two distributed algorithms were devised to resolve the problem. It is noteworthy that, to our best knowledge, [21] is the first to investigate the fixed-point finding problem of a global operator in real Hilbert spaces, where the global operator is an average of local operators over a multiagent network. Nevertheless, the convergence rate is not analyzed for the proposed algorithms in [21].

Motivated by the above facts, the purpose of this article is to further investigate the fixed-point finding problem of a quasi-nonexpansive global operator over a time-invariant, directed, and unbalanced communication graph. Two scenarios are taken into account, i.e., the global operator is sum separable and block separable. In the first case, the global operator is composed of a sum of local operators, and in the second case, the global operator is comprised of a family of local block mappings. In both cases, local operators/mappings are assumed to be Lipschitz and are only privately accessible to each individual agent, thereby needing all agents to tackle the global problem in a collaborative manner. The contributions of this article are threefold as follows.

1) For the first case, a distributed algorithm, called distributed quasi-averaged operator tracking algorithm (DOT), is developed and shown to be convergent to a fixed point of the global operator at a linear rate under a linear regularity condition. Compared with the closely related work [21], where no convergence speed is provided, a different algorithm is developed here and shown to be convergent at a linear rate. It should be noted that the problem here is more general than the common fixed-point seeking problem [16], [17], [18], [20], where all local operators are assumed to have at least one common fixed point, while this assumption is dropped here. As a special case, linear convergence can also be ensured for the common fixed-point seeking problem. In contrast, [20] only provided a sublinear rate for nonexpansive operators.

2) For the second case, a distributed algorithm, called distributed quasi-averaged operator playing algorithm (DOP), is proposed, which is shown to be linearly convergent to a fixed point of the global operator under the linear regularity condition. To our best knowledge, this is the first to study the block separable case in a decentralized manner.

3) The studied setups in this article provide a unified framework for a host of interesting problems. For example, the proposed DOT and DOP algorithms can be leveraged to resolve distributed optimization and multiplayer games under partial-decision information.

A preliminary version of this article was presented at a conference [27]. The present article extends the result of [27] in several ways. The authors in [27] only considered the sum separable case without providing detailed proofs of the main result. In comparison, a full proof of the main result (i.e., Theorem 1) for the sum separable case is provided. Besides, one more scenario is investigated in this article, i.e., the block separable case (see Theorem 2). Also, one more application is presented here, i.e., multiplayer games under partial-decision information, along with one more numerical example.

The rest of this article is organized as follows. Some basic knowledge and the problem formulation are introduced in Section II, and the first case with the global operator being sum separable is addressed in Section III, followed by the second case with the global operator being block-coordinate separable in Section IV. Several applications are provided in Section V. Some numerical examples are presented in Section VI. Finally, Section VII concludes the article.

II. PRELIMINARIES AND PROBLEM FORMULATION

A. Notations

Let $\mathcal{H}$ be a real Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and associated norm $\| \cdot \|$. Note that $\mathcal{H}$ can be both finite and infinite dimensional, including $\mathbb{R}^n$ as a special case. Define $[N] := \{1, 2, \ldots, N\}$ for any integer $N > 0$, and denote by $\text{col}(z_1, \ldots, z_k)$ the column vector or matrix by stacking up $z_i, i \in [k]$. Given an integer $n > 0$, denote by $\mathbb{R}, \mathbb{R}^n, \mathbb{R}^{n \times n}$, and $\mathbb{N}$ the sets of real numbers, $n$-dimensional real vectors, $n \times n$ real matrices, and nonnegative integers, respectively. Let $P_X(\cdot)$ represent the projection operator onto a nonempty closed
and convex set $X \subseteq \mathcal{H}$, i.e., $P_X(z) := \arg\min_{x \in X} \|z - x\|$ for $z \in \mathcal{H}$. Moreover, denote by $I$, $I_d$, and $\otimes$ the identity matrix of appropriate dimension, the identity operator, and the Kronecker product, respectively. Let $I_n$ be an $n$-dimensional vector with all entries 1 for an integer $n > 0$, and the subscript is omitted when the dimension is clear in the context. $d_X(z) := \inf_{x \in X} \|z - x\|$ denotes the distance from $z \in \mathcal{H}$ to the set $X$. For an operator $T : \mathcal{H} \to \mathcal{H}$, define $\text{Fix}(T) := \{x \in \mathcal{H} | T(x) = x\}$ to be the set of fixed points of $T$ and $T_\beta := I + \beta(T - I_d)$, called a $\beta$-relaxation of $T$ with a relaxation parameter $\beta \geq 0$. Denote by $M_\infty$ the infinite power of a square matrix $M$, i.e., $M_\infty = \lim_{k \to \infty} M^k$, if it exists, and let $\rho(M)$ and $\det(M)$ be the spectral radius and determinant of $M$, respectively.

**B. Operator Theory**

Consider an operator $T : S \to \mathcal{H}$ for a nonempty set $S \subseteq \mathcal{H}$. $T$ is called $L$-Lipschitz (continuous) for a constant $L > 0$ if

$$\|T(x) - T(y)\| \leq L \|x - y\|, \quad \forall x, y \in S. \quad (1)$$

Further, $T$ is called nonexpansive (resp. contractive) if $L = 1$ (respectively, $L < 1$), quasi-nonexpansive (QNE) if $(1)$ holds with $L = 1$ for all $x \in S$ and $y \in \text{Fix}(T)$, and $\rho$-strongly quasi-nonexpansive ($\rho$-SQNE) for $\rho > 0$ if it holds that for all $x \in S$ and $y \in \text{Fix}(T)$

$$\|T(x) - y\|^2 \leq \|x - y\|^2 - \rho \|x - T(x)\|^2. \quad (2)$$

$T$ is called $\eta$-averaged (resp. $\eta$-quasi-averaged) for $\eta \in (0, 1)$ if it can be written as

$$T = (1 - \eta)I_d + \eta R \quad (3)$$

for some nonexpansive (resp. quasi-nonexpansive) operator $R$.

The above concepts can be found in [1, Section 4.1], where quasi-averaged operators are defined here as an analogue of averaged operators. It is well known that when $T$ is QNE, the set $\text{Fix}(T)$ is closed and convex (cf., Corollary 4.24 in [1]).

The operator $T$ is said linearly regular [28] if there exists a constant $\omega > 0$ such that

$$d_{\text{Fix}(T)}(x) \leq \omega \|x - T x\|, \quad \forall x \in S.$$  

It is easy to observe that the linear regularity means that the distance between $x$ and $T x$ is lower bounded by a scaled distance from the vector $x$ to the set $\text{Fix}(T)$. For example, the projection operator $P_C$ is linearly regular with constant 1, where $C \subseteq \mathcal{H}$ is nonempty, closed, and convex [29]. Moreover, it has been shown in [30] that for an SQNE operator, linear regularity is necessary and sufficient for global Q-linear convergence to a fixed point with rate independent of the starting point. More details on linear regularity can be found in [29], [30], and [31].

**C. Graph Theory**

The communication pattern among all agents is captured by a simple directed graph, denoted by $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = [N]$ and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ are the node (or agent) and edge sets, respectively. An edge $(i, j) \in \mathcal{E}$ means that agent $i$ can send information to agent $j$, but not necessarily vice versa, and agent $i$ (resp. $j$) is called an in-neighbor or simply neighbor (respectively, out-neighbor) of agent $j$ (respectively, $i$). A graph is called undirected, if and only if $(i, j) \in \mathcal{E}$ amounts to $(j, i) \in \mathcal{E}$, and directed otherwise. A directed path is defined to be a sequence of adjacent edges $(i_1, i_2), (i_2, i_3), \ldots, (i_{n-1}, i_n)$, and a graph is said strongly connected if any two nodes can be connected by a directed path from one to the other.

**D. Problem Formulation**

The aim of this article is to compute a fixed point of a global operator $F : \mathcal{H} \to \mathcal{H}$, i.e.,

$$\text{find } x \in \mathcal{H}, \text{ s.t. } x \in \text{Fix}(F). \quad (4)$$

It is worth mentioning that application examples of problem (4) in real Hilbert spaces, but not in $\mathbb{R}^n$, include digital signal processing and $L^2([0, \pi])$ (i.e., the space of all square integrable functions $f : [0, \pi] \to \mathbb{R}$) [32], and so on.

In this article, no global/central coordinator, master, or computing unit is assumed to exist for problem (4), and however the global operator $F$ is separable and consists of local operators, which are privately accessible to each individual agent in a network. Briefly speaking, only partial information of $F$ can be privately known by each agent in a network, which is interesting and realistic in large-scale problems, as extensively studied in distributed optimization and distributed machine learning, and so on. Specifically, we focus on two scenarios: 1) $F$ is sum separable; and 2) $F$ is block separable, as elaborated below.

**Case 1. Sum Separable:** $F$ is sum separable over a network of $N$ agents, i.e.,

$$\text{(Problem I)} \quad \text{find } x \in \text{Fix}(F), \quad F = \frac{1}{N} \sum_{i=1}^{N} F_i \quad (5)$$

where each $F_i : \mathcal{H} \to \mathcal{H}$ is a local operator, only privately accessible to agent $i$ for $i \in [N]$. Note that the formulation (5) is also investigated in [21], where, however, no convergence rates are provided, while a linear convergence speed is established in this article. The formulation (6) is more general than the common fixed-point finding problem [16], [17], [18], [19], [20], [21] and the linear algebraic equation solving problem [24], [25], [26], as discussed in the introduction section.

**Case 2. Block Separable:** In this case, $\mathcal{H} = \mathcal{H}_1 \oplus \cdots \oplus \mathcal{H}_N$ is the direct Hilbert sum, where every $\mathcal{H}_i, i \in [N]$ is a real Hilbert space, and the inner products and associated norms of these spaces are all denoted as $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$, respectively.

Let $x = (x_1, \ldots, x_N)$ denote a generic vector in $\mathcal{H}$ with $x_i \in \mathcal{H}_i, i \in [N]$. Then the global operator $F$ can be written as a block-coordinate version $F = (F_1, \ldots, F_N)$, where $F_i : \mathcal{H} \to \mathcal{H}_i$, for $i \in [N]$, i.e., $F(x) = (F_1(x_1), \ldots, F_N(x_N))$ for $x \in \mathcal{H}$. In this setup, $F$ is block (or block-coordinate) separable over a network of $N$ agents, i.e.,

$$\text{(Problem II)} \quad \text{find } x \in \text{Fix}(F), \quad F = (F_1, \ldots, F_N) \quad (6)$$

where each $F_i$ is only privately accessible to agent $i$ who only knows its own vector $x_i$ with no knowledge of $x_j$’s for all $j \neq i \in [N]$. To our best knowledge, this scenario is novel and also practical (c.f. an application to multiplayer games in Section V-B).
With the above discussion, the objective of this article is to develop distributed algorithms to solve Problems I and II.

Remark 1: To illustrate applications of the above problems, a simple example in function approximation is provided here, which is useful such as in reinforcement learning (cf., Chapter 9 in [34]). Consider a (reward) function \( r: \mathbb{R}^n \rightarrow \mathbb{R} \), which may be unknown in reality, and let us approximate it by \( \sum_{j=1}^{\infty} c_j \exp(-\frac{|x-c_j|^2}{2\sigma_j^2}) \) based on radial basis functions (RBFs), where \( c_j \) and \( \sigma_j \) are some prespecified parameters (e.g., feature’s center state and feature’s width in reinforcement learning, respectively), \( w = (w_1, w_2, \ldots) \in \ell^2 \) is the variable to be optimized, and \( \ell^2 \) is the space of square-summable sequences which is an infinite-dimensional Hilbert space. Then, the performance is to minimize the approximation error \( f(w) := \frac{1}{N} \sum_{i=1}^{N} f_i(w) \) with \( f_i(w) := \frac{1}{N} \sum_{s_i} |r(s_i) - \sum_{j=1}^{\infty} w_j \exp(-\frac{|s_i-c_j|^2}{2\sigma_j^2})|^2 \), where \( \{s_i\}_{i=1}^{N} \) are a set of sample data privately known by agent \( i \). It is easy to verify that \( f_i \) is differentiable and convex with \( L_i \)-Lipschitz gradient for some constant \( L_i > 0 \), thus implying that \( F_i: w \rightarrow w - \xi \nabla f_i(w) \) is non-expansive for a constant \( \xi \in (0, 2/L) \) with \( L := \max_{i \in [N]} L_i \) (3, Lemma 4). The problem then amounts to the fixed-point finding problem (5) with \( \mathcal{H} = \ell^2 \). More applications will be given in Section V.

E. Assumptions

With the above preparations, we are now ready to impose some standard assumptions.

Assumption 1 (Strong Connectivity): The communication graph \( \mathcal{G} \) is strongly connected. Moreover

1) Two matrices \( A = (a_{ij}) \in \mathbb{R}^{N \times N} \) (row-stochastic) and \( B = (b_{ij}) \in \mathbb{R}^{N \times N} \) (column-stochastic) are arbitrarily assigned to \( \mathcal{G} \) with \( a_{ii} > 0, b_{ii} > 0 \) for all \( i \in [N] \).

2) Denote the left stochastic eigenvector (respectively, right stochastic eigenvector) of \( A \) (respectively, \( B \)) associated with the eigenvalue \( 1 \) as \( \pi = (\pi_1, \ldots, \pi_N) \) (respectively, \( \nu = (\nu_1, \ldots, \nu_N) \)) such that \( A_{\infty} = 1 \pi \nu^\top \) and \( \pi_1 > 0 \) (respectively, \( B_{\infty} = \nu 1^\top \pi \) and \( \nu_i > 0 \) for all \( i \in [N] \)).

It is worth mentioning that the directed graph is not required to be balanced in this article. Note that, similar to [35], \( A \) and \( B \) are consistent with \( \mathcal{G} \), i.e., \( a_{ij} > 0 \) and \( b_{ij} > 0 \), if and only if \( (j,i) \in \mathcal{E} \) for \( i \neq j \). Notice that \( A \) and \( B \) do not need to be doubly stochastic. Additionally, the property in Assumption 1.2 can be ensured by the strong connectivity of \( \mathcal{G} \) [35].

Assumption 2: \( F_i \) is Lipschitz with constant \( L_i \) for all \( i \in [N] \), i.e., \( \|F_i(x) - F_i(y)\| \leq L_i \|x - y\|, \forall x, y \in \mathcal{H} \).

Assumption 3:

1) \( F \) is quasi-nonexpansive with \( F\text{fix}(F) \neq \emptyset \).

2) \( F \) is linearly regular, i.e., there exists a constant \( \kappa > 0 \) such that \( d_{F\text{fix}(F)}(x) \leq \kappa \|F(x) - x\|, \forall x \in \mathcal{H} \). (7)

Remark 2: Note that Assumption 3 is only made for the global operator \( F \), and is not necessary for any local operators \( F_i \)’s.

III. DOT Algorithm for Problem I

This section aims to develop a distributed algorithm for tackling problem (5) which can converge at a linear rate. Without loss of generality, the vectors in \( \mathcal{H} \) are viewed as column vectors in this section.

For problem (5), if \( F \) can be known by a global/central computing unit (or coordinator), then a famous centralized algorithm, called the KM iteration [36], can be exploited, i.e.,

\[
x_{k+1} = x_k + \alpha_k (F(x_k) - x_k)
\]  

(8)

where \( \{\alpha_k\}_{k \in \mathbb{N}} \) is a sequence of relaxation parameters with \( \alpha_k \in [0,1] \). Note that the KM iteration usually applies to non-expansive operators, but it still works for quasi-nonexpansive operators here under the linear regularity condition in Assumption 3. However, the centralized iteration (8) is not realistic here since no global/central computing unit (or coordinator) exists in our setting, which hence motivates us to devise distributed (or decentralized) algorithms based on only local information exchanges among all agents.

Motivated by the classical KM iteration and the tracking techniques such as those in [35], [37], [38], a distributed quasi-averaged operator tracking algorithm (DOT) is proposed as follows:

\[
x_{i,k+1} = \sum_{j=1}^{N} a_{ij} x_{j,k} + \alpha \left( \frac{y_{i,k}}{w_{i,k}} - \sum_{j=1}^{N} a_{ij} x_{j,k} \right),
\]

(9a)

\[
y_{i,k+1} = \sum_{j=1}^{N} b_{ij} y_{j,k} + F_i(x_{i,k+1}) - F_i(x_{i,k}),
\]

(9b)

\[
w_{i,k+1} = \sum_{j=1}^{N} b_{ij} w_{j,k},
\]

(9c)

where \( x_{i,k} \) is an estimate of a fixed point of the global operator \( F \) by agent \( i \) at time \( k \geq 0 \) for all \( i \in [N] \), and \( \alpha \in (0,1) \) is the stepsize to be determined. Set the initial conditions as: arbitrary \( x_{i,0} \in \mathcal{H}, y_{i,0} = F_i(x_{i,0}) \), and \( w_{i,0} = 1 \) for all \( i \in [N] \). It is noteworthy that neighboring agents are only involved in (9) for each agent due to \( a_{ij} = 0 \) and \( b_{ij} = 0 \) when agent \( j \) is not a neighbor of agent \( i \).

Roughly speaking, \( y_{i,k} \) is employed to track the weighted global operator \( \nu_i \sum_{i=1}^{N} F_i(x_{i,k}) \), and meanwhile \( w_{i,k} \) is a scalar used to track \( \nu_i N \) in order to counteract the imbalance of the matrix \( B \) in (9b).

For (9c), it is easy to verify that each \( w_{i,k} \) will exponentially converge to \( N \nu_i \). However, invoking the method in [39], it can be concluded that the final value \( N \nu_i \) can be evaluated for each agent \( i \) in finite time in a distributed manner. Because of this, without loss of generality, algorithm (9) can be rewritten by replacing \( w_{i,k} \) with \( N \nu_i \) as in Algorithm 1.

To facilitate the following analysis, Algorithm 1 can be written in a compact form:

\[
x_{k+1} = A x_k + \alpha \left( \frac{1}{N} (D \nu^{-1} \otimes I d) y_k - A x_k \right),
\]

(10)

\[
y_{k+1} = B y_k + F(x_{k+1}) - F(x_k)
\]

(11)

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
where $x_k, y_k$ are concatenated vectors of $x_i, y_i$, respectively, $A := A \otimes I_d$, $B := B \otimes I_d$, $D_o := \text{diag}(\nu_1, \ldots, \nu_N)$, and $F(z) := \text{col}(F_1(z_1), \ldots, F_N(z_N))$ for a vector $z = \text{col}(z_1, \ldots, z_N) \in \mathcal{H}^N := \mathcal{H} \times \cdots \times \mathcal{H}$ (the $N$-fold Cartesian product of $\mathcal{H}$).

To proceed, it is helpful to introduce two new weighted norms for the Cartesian product $\mathcal{H}^N$, i.e.,

$$
\|z\|_\pi := \sqrt{\sum_{i=1}^{N} \pi_i \|z_i\|^2}, \quad \|z\|_\nu := \sqrt{\sum_{i=1}^{N} \nu_i \|z_i\|^2} \quad (13)
$$

for any vector $z = \text{col}(z_1, \ldots, z_N) \in \mathcal{H}^N$. Let $\| \cdot \|$ be the natural norm in $\mathcal{H}^N$, i.e., $\|z\| := \sqrt{\sum_{i=1}^{N} \|z_i\|^2}$. Additionally, it is also necessary to introduce two weighted norms in $\mathbb{R}^N$ [35], i.e., for any $x = \text{col}(x_1, \ldots, x_N) \in \mathbb{R}^N$

$$
\|x\|_\pi := \sqrt{\sum_{i=1}^{N} \pi_i x_i^2}, \quad \|x\|_\nu := \sqrt{\sum_{i=1}^{N} \nu_i x_i^2}. \quad (14)
$$

Please note that the notations $\| \cdot \|_\pi$ and $\| \cdot \|_\nu$ in (13) and (14) should be easily distinguished by the context. Accordingly, let us denote by $\|M\|_\pi$ and $\|M\|_\nu$ (resp. $\|M \otimes T\|_\pi$ and $\|M \otimes T\|_\nu$) the norms for a matrix $M \in \mathbb{R}^{N \times N}$ (respectively, a matrix $M \in \mathbb{R}^{N \times N}$ and an operator $T$ in $\mathcal{H}$) induced by $\| \cdot \|_\pi$ and $\| \cdot \|_\nu$ in (14) (respectively, (13)), respectively.

It is easy to see that the natural norm $\| \cdot \|$ is equivalent to $\| \cdot \|_\pi$, $\| \cdot \|_\nu$ in (13), (14), and thus to the induced matrix norms, that is, there are positive constants $c_1, \ldots, c_4$ in [4] such that

$$
c_1 \cdot \| \cdot \|_\pi \leq \| \cdot \| \leq c_2 \cdot \| \cdot \|_\pi, \quad (15)
$$

$$
c_3 \cdot \| \cdot \|_\nu \leq \| \cdot \| \leq c_4 \cdot \| \cdot \|_\nu. \quad (16)
$$

Then, the following results can be obtained.

**Lemma 1 ([35]):** For all $x \in \mathbb{R}^N$, there hold

$$
\|Ax - A_\infty x\|_\pi \leq \rho_1 \|x - A_\infty x\|_\pi, \quad (17)
$$

$$
\|Bx - B_\infty x\|_\nu \leq \rho_2 \|x - B_\infty x\|_\nu, \quad (18)
$$

$$
\|A\|_\pi = \|A_\infty\|_\pi = \|I_N - A_\infty\|_\pi = 1, \quad (19)
$$

$$
\|B\|_\nu = \|B_\infty\|_\nu = \|I_N - B_\infty\|_\nu = 1. \quad (20)
$$

where $\rho_1 := \|A - A_\infty\|_\pi < 1$ and $\rho_2 := \|B - B_\infty\|_\nu < 1$.

**Lemma 2:** For all $x \in \mathcal{H}^N$, the following statements hold:

$$
\|Az - A_\infty z\|_\pi \leq \rho_1 \|z - A_\infty z\|_\pi, \quad (21)
$$

$$
\|Bz - B_\infty z\|_\nu \leq \rho_2 \|z - B_\infty z\|_\nu, \quad (22)
$$

$$
\|I_N \otimes Id - A_\infty\|_\pi = \|I_N - A_\infty\|_\pi = 1 \quad (23)
$$

where $A_\infty := I_N \otimes Id$, $B_\infty := B \otimes Id$, and $\rho_1, \rho_2$ are defined in Lemma 1.

**Proof:** The proof can be found in Appendix A.

**Lemma 3:** It holds that $\bar{y}_k = \sum_{i=1}^{N} F_i(x_i, k)$, where $\bar{y}_k := \sum_{i=1}^{N} y_k$.

**Proof:** Left multiplying (11) by $1^\top$ yields that $\bar{y}_{k+1} = \bar{y}_k + \sum_{i=1}^{N} F_i(x_i, k+1) - \sum_{i=1}^{N} F_i(x_i, k)$, which further implies that $\bar{y}_k - \sum_{i=1}^{N} F_i(x_i, k) = \bar{y}_0 - \sum_{i=1}^{N} F_i(x_i, 0)$. Note that $\bar{y}_0 = F_1(x_{i, 0})$. The conclusion directly follows.

To move forward, an important result for the convergence analysis is first given below.

**Lemma 4:** Under Assumption 3, if $\alpha \in (0, 1 - \delta)$, then for any vector $x \in \mathcal{H}$, there holds

$$
d_{F_i(x)}(F_\alpha(x)) \leq \rho_3 d_{F_i(x)}(x), \quad \forall x \in \mathcal{H} \quad (24)
$$

where $F_\alpha := Id + \alpha(F - Id)$ and

$$
\rho_3 := 1 - \frac{\delta \alpha}{2\kappa^2} \in [0, 1]. \quad (25)
$$

**Proof:** It is easy to see that $F_\alpha - Id = \alpha(F - Id)$, which together with (7) yields that $d_{F_i(x)}(x) \leq \|F_i(x) - x\| = \frac{\delta \alpha}{\kappa} \|F_i(x) - x\|$ for all $x \in \mathcal{H}$. Therefore, $F_i(x)$ is linearly regular with constant $\frac{\delta \alpha}{\kappa}$. Simultaneously, it is known that each $\alpha$-quasi-averaged operator is $\frac{1 - \alpha}{\alpha}$-SQNE [30], and thus $F_\alpha$ is $\frac{1}{\alpha}$-SQNE. With the above two properties of $F_\alpha$ as well as $F_i(x) = F_i(x_i, k)$, invoking Theorem 1 in [30] leads to

$$
d_{F_i(x)}(F_\alpha(x)) \leq \phi d_{F_i(x)}(x), \quad \forall x \in \mathcal{H} \quad (26)
$$

where $\phi := \sqrt{1 - \frac{\alpha(1 - \alpha)}{2\kappa^2}} \in (0, 1)$. Meanwhile, it is easy to verify that

$$
\phi \leq 1 - \frac{\alpha(1 - \alpha)}{2\kappa^2} \leq 1 - \frac{\delta \alpha}{2\kappa^2}
$$

where $\alpha \leq 1 - \delta$ is used in the last inequality. This ends the proof.

**Theorem 1:** Under Assumptions 1–3, the quantity $\|x_{i,k} - x_k\|$ converges to 0 at a linear rate for all $i \in [N]$, where $x_k := P_{F_i(x)}(\bar{x}_k)$ and $\bar{x}_k := \sum_{i=1}^{N} \pi_i x_{i,k}$, if there holds

$$
0 < \eta < \min\{1 - \delta, \alpha_c\} \quad (27)
$$

where $\alpha_c$ is the smallest positive real root of equation $\det(I - M(\alpha)) = 0$, and

$$
M(\alpha) := \begin{pmatrix}
\alpha_1 & \alpha_2 & \cdots & 0 \\
\alpha_2 & \alpha_3 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & \cdots & \cdots & \alpha_N
\end{pmatrix}
$$

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
Algorithm 2: Distributed Quasi-Averaged Operator Playing (DOP).

1: **Initialization:** Stepsize $\alpha$ in (32), communication matrix $A$, and local initial conditions $x_1^0 \in \mathcal{H}$ for all $i \in [N]$.

2: **Iterations:** Step $k \geq 0$: update for each $i \in [N]$:

\[
x_{i,k+1} = \sum_{j=1}^{N} a_{ij} x_{i,k}^j + \frac{\alpha}{\pi_i} \left( F_i(x_{i,k}^j) - \sum_{j=1}^{N} a_{ij} x_{i,k}^j \right),
\]

\[
x_{i,k+1} = \sum_{j=1}^{N} a_{ij} x_{j,i,k}^j.
\]

IV. DOP ALGORITHM FOR PROBLEM II

This section is concerned with solving problem (6). Without loss of generality, the vectors in $\mathcal{H}$ are viewed as row vectors in this section for convenient analysis.

For problem (6), each agent $i \in [N]$ can only privately access $F_i$ with its own vector $x_i$ for a whole vector $x = (x_1, \ldots, x_N) \in \mathcal{H}$ over a network of $N$ agents, where $x_i$ is privately known by agent $i$ itself, as commonly encountered in multiplayer games, and so on. To handle this problem, each agent $i \in [N]$ maintains a vector $x_i = (x_{i,1}, \ldots, x_{i,N}) \in \mathcal{H}$ at time step $k \geq 0$ as an estimate of a fixed point of $F$, where $x_{i,k}$ is an estimate of $x_{i,k}$ (i.e., the vector of agent $j$ at time $k$) by agent $i$ at time $k$ with $x_{i,k} = x_{i,k}$. That is, each agent $i$ updates its own vector $x_{i,k}$ at time $k$ without access to the vectors of all other agents $j \neq i$, and thus each agent $i$ needs to estimate other agents’ vectors $x_{j,k}$ denoted as $x_{j,k}$ at time $k$ over the communication graph $G$ satisfying Assumption 1.

Now, a distributed algorithm is proposed as in Algorithm 2, where $A = (a_{ij}) \in \mathbb{R}^{N \times N}$ is the communication matrix introduced after Assumption 1, which is only row-stochastic, and $x_{j,i,k} := (x_{j,1,k}, \ldots, x_{j,1,k}^j, x_{j,i,k}^j, \ldots, x_{j,N,k}^j)$ for all $i, j \in [N]$, i.e., $x_{j,i,k}$ is the agent $j$’s estimate of all agents’ vectors except the $i$th agent. We recall that $\pi$ is the left stochastic eigenvector of $A$ associated with the eigenvalue 1 as introduced in Assumption 1.2. It should be noted that the $i$th entry $\pi_i$ of $\pi$ can be evaluated by agent $i$ in finite time in a distributed fashion using the approach in [39]. Thus, Algorithm 2 is distributed.

To ease the upcoming analysis, let us define $x_k := \operatorname{col}(x_1^0, \ldots, x_N^0) \in \mathcal{H}^N$, $\tilde{x}_{i,k} := \sum_{j=1}^{N} a_{ij} x_{i,k}^j$, and $\tilde{F} := \operatorname{diag}(\{F_1(x_1^0) - \tilde{x}_1, \ldots, (F_N(x_N^0) - \tilde{x}_N)/\pi \})$. Then algorithm (28) can be written in a compact form

\[
x_{k+1} = Ax_k + \alpha \tilde{F}.
\]

Multiplying $\pi^\top$ on both sides of (29) yields that

\[
\tilde{x}_{k+1} = \tilde{x}_k + \alpha \tilde{F}.
\]

where $\tilde{x}_k = (\tilde{x}_1, \ldots, \tilde{x}_N)$ and $\tilde{F} := (F_1(x_1^0) - \tilde{x}_1, \ldots, F_N(x_N^0) - \tilde{x}_N)$.
To move forward, it is useful to recall the weighted norm
\[ \|z\|_\pi := \sqrt{\sum_{i=1}^{N} \pi_i \|z_i\|^2} \]
for a vector \( z = \text{col}(z_1, \ldots, z_N) \in \mathcal{H}^N \), as defined in (13). And let \( \| \cdot \| \) be a norm in \( \mathcal{H}^N \) defined by
\[ \|z\| := \sqrt{\sum_{i=1}^{N} \|z_i\|^2}. \]
Remember that the vectors in \( \mathcal{H} \) are seen as row vectors in this section. Then, similar to (21) in Lemma 2, it is easy to obtain the following result.

**Lemma 5:** For all \( z \in \mathcal{H}^N \), it holds that
\[ \|Ax - A_{\infty}z\|_\pi \leq \rho_1 \|z - A_{\infty}z\|_\pi \]  
(31)
where \( A_{\infty} = 1_N \pi^T \) as defined in the paragraph after Assumption 1 and \( \rho_1 := \|A - A_{\infty}\|_\pi < 1 \).

To ensure the linear convergence, Assumptions 2 and 3 are still imposed here, but \( F_i \) in Assumption 2 is replaced with \( F_i \), i.e., \( \|F_i(x) - F_i(y)\| \leq L_i \|x - y\| \), \( \forall x, y \in \mathcal{H} \) for \( i \in [N] \).

With the above preparations, we are now ready to give the main result of this section.

**Theorem 2:** Under Assumptions 1–3 with \( F_i \) being replaced with \( F_i \) in Assumption 2, the quantity \( \|x_k^i - \bar{x}_k\| \) converges to 0 at a linear rate for all \( i \in [N] \), where \( \bar{x}_k = \sum_{i=1}^{N} \pi_i x_k^i \), if
\[ 0 < \alpha < \min\{1 - \delta, \alpha_L\} \]  
(32)
where \( \delta \in (0, 1) \) is any prespecified parameter, \( \alpha_L \) is the smallest positive real root of \( \det(I - \Theta(\alpha)) = 0 \), and
\[ \Theta(\alpha) := \left( \frac{\rho_1 + \alpha \theta_5}{\alpha(L+1)} \right) \frac{2\alpha c_2 \sqrt{\bar{v}}}{1 - \frac{\delta_0}{\bar{v}}^2} \]  
(33)
with \( \theta_5 := 2c_2 \sqrt{\bar{v}(L^2 + 1)/\epsilon_1} \), \( \bar{L} := \max_{i \in [N]} \{L_i\} \), \( \bar{v} := N - 1 + \frac{(1-\pi)^2}{2} \), and \( \bar{\pi} := \min_{i \in [N]} \{\pi_i\} > 0 \).

**Proof:** The proof can be found in Appendix C. \( \square \)

**Remark 4:** Note that \( \rho(\Theta(\alpha)) < 1 \) under (32). It is also noteworthy that the work [21] only considers the sum separable case and does not present the convergence speed. In contrast, this article addresses both the sum separable case (see Theorem 1) and the block separable case (see Theorem 2) with linear rates, and to our best knowledge, this article is the first to address the block separable case for the fixed-point finding problem in the decentralized fashion.

**Remark 5:** It is worth noting that Problem II can be cast as the common fixed-point finding problem for a family of operators \( T_i := (\text{Id}_1, \ldots, \text{Id}_{i-1}, F_i, \text{Id}_{i+1}, \ldots, \text{Id}_N) : \mathcal{H} \to \mathcal{H} \), where \( \text{Id}_i : x \mapsto x_i \) for \( x = (x_1, \ldots, x_N) \in \mathcal{H} \) and \( i \in [N] \).

Define an operator \( T := \frac{1}{N} \sum_{i=1}^{N} T_i \). It is easy to verify that \( T = \frac{N-1}{N} \text{Id} + \frac{1}{N} F \) and \( T \) is linearly regular with constant \( N \kappa \).

Therefore, Algorithm 1 can be leveraged to solve Problem II at a linear rate by Corollary 1. However, compared with Algorithm 2, there are three shortcomings when doing so as follows.

1) In Algorithm 1, two \( \mathcal{H} \)-dimensional vectors \( x_{i,k} \) and \( y_{i,k} \) are maintained, updated, and transmitted for each agent \( i \in [N] \), which may incur high burdens of computation, storage, and communication especially for large dimensional spaces. In contrast, by using Algorithm 2, only one \( \mathcal{H} \)-dimensional vector \( x_{i,k} \) is maintained, updated, and transmitted for each agent \( i \).

2) Algorithm 1 depends on two communication matrices \( A \) and \( B \), which are row- and column-stochastic, respectively, while Algorithm 2 builds upon only one row-stochastic matrix \( A \), which is easier to implement.

3) With the linear regularity constant being \( N \kappa \) for \( T \), the \( 3 \times 3 \) element of \( M(\alpha) \) in (27) will equal \( 1 - \frac{\delta_0}{N^2 \bar{v}^2} \), which will generally narrow down the feasible range of stepsize \( \alpha \) in (26). However, directly using the linear regularity constant \( \kappa \) of \( F \) in Algorithm 2, the last entry \( 1 - \frac{\delta_0}{\bar{v}^2} \) in (33) is independent of \( N \), potentially leading to a larger feasible range for \( \alpha \).

**Remark 6:** It is noteworthy that in problem II each agent \( i \) can only know its own vector \( x_{i,k} \) at each time \( k \geq 0 \), but has no access to vectors \( x_{j,k} \)'s of all other agents for \( j \neq i \). In this regard, agent \( i \) needs to estimate all other \( x_{j,k} \)'s in order to compute the value of its \( F_i \). If each agent has full access to all other agents’ vectors, then a simpler algorithm can be devised to tackle this setup, i.e.,
\[ x_{i,k+1} = x_{i,k} + \alpha(F_i(x_k) - x_{i,k}) \]  
(34)
where \( x_{i,k} \) is the same as in ((28)) and \( x_k := (x_{1,k}, \ldots, x_{N,k}) \).

However, there is no need for each agent to estimate the entire vector \( x_k \) in this setup. As for (34), a linear convergence to a fixed point of \( F \) can be similarly obtained to Theorem 2.

V. APPLICATIONS OF DOT AND DOP

The studied problems can provide a unified framework for a multitude of interesting problems. To show this, this section aims to provide two examples, i.e., distributed optimization and multiplayer games under partial-information decision.

A. Distributed Optimization

Consider a global optimization problem
\[ \min_{x \in \mathcal{H}} f(x) \]  
(35)
where \( f : \mathcal{H} \to \mathbb{R} \) is a differentiable and convex function, whose gradient is Lipschitz with constant \( L \). It is easy to verify that problem is equivalent to finding fixed points of an operator \( F : x \mapsto x - \xi \nabla f(x) \) for any given \( \xi > 0 \), which is shown to be \((L\xi)/2\)-averaged when \( \xi \in (0, 2/L) \) (cf. [3, Lemma 4]) and thus nonexpansive. For many realistic problems, it is usually difficult or impossible to access \( f \) by a global/central coordinator or computing unit, instead it is more practical for \( f \) to be separable. In instance, in distributed estimation of a parameter (e.g., a robot’s position) by a collection of wireless sensors that are dislocated over large regions, each sensor is only capable of measuring partial information on the parameter according to its own position and sensing ability, and to evaluate the parameter as precisely as possible, it is necessary to collect all the sensors’ information, which, however, is not known by any single sensor [40, 41]. Along this line, two cases are discussed below.

**Case 1:** \( f \) is sum separable, i.e., \( f(x) = \frac{1}{N} \sum_{i=1}^{N} f_i(x) \), where \( f_i : \mathcal{H} \to \mathbb{R} \) is the local function, which is differentiable and convex with \( L_i \)-Lipschitz gradient, only known to agent \( i \).
This problem is often called distributed/decentralized optimization, which has been extensively studied in the literature. In this case, the problem can be equivalently cast as problem I (i.e., (5)) with $F_i : x \mapsto x - \xi \nabla f_i(x)$ for any given bounded $\xi > 0$, which is Lipschitz. Therefore, Assumption 2 holds true. In this setup, the problem can be recast as problem (6) with $F_i : x \mapsto x - \xi \nabla f_i(x)$ for any given bounded $\xi > 0$, which is Lipschitz if $\nabla f_i (x)$ is so. For this problem, under Assumptions 1–3, the linear convergence to a solution of (35) can be guaranteed by Theorem 1.

Case 2: $f$ is block separable, i.e., $\nabla f(x) = \text{col}(\nabla_{x_1} f(x), \ldots, \nabla_{x_N} f(x))$ with $x = \text{col}(x_1, \ldots, x_N)$, where $x_i$ is the vector of agent $i \in [N]$ and each agent $i$ is only capable of computing partial gradient $\nabla_{x_i} f(x)$ with respect to its own vector $x_i$. This scenario is realistic in some cases, partially because it is computationally expensive to compute the whole gradient $\nabla f(x)$ by a global/central coordinator or computing unit, and partially because only part of data $x_i$ may be privately acquired by spatially distributed agents. In this setup, the problem can be recast as problem (6) with $F_i : x \mapsto x - \xi \nabla f_i(x)$ for any given bounded $\xi > 0$, which is Lipschitz if $\nabla f_i(x)$ is so. For this problem, under Assumptions 1–3, the linear convergence to a solution of (35) can be guaranteed by Theorem 2.

Remark 7: Note that in Case 1, the linear convergence rate is ensured under the linear regularity of $F$ in Assumption 3, which is strictly weaker than the strong convexity of $f_i$’s or $f$ (cf. Section III in [30]), widely postulated in distributed optimization (e.g., [40], [41], [42], [43], [44], [45]). That is, the strong convexity of smooth $f_i$’s or $f$ can imply the linear regularity of $F$, but not vice versa. Also, notice that the linear regularity is only assumed for $F$, not necessary for any local $F_i$. For Case 1, a similar condition to linear regularity, i.e., metric subregularity, is employed in [46] for deriving a linear convergence, which is, however, in Euclidean spaces under balanced undirected communication graphs, while the result here works in a more general setting, i.e., in real Hilbert spaces under unbalanced directed graphs. It should be also noted that the aforesaid problem is just an application of a general problem (5) addressed here. In addition, to our best knowledge, this article is the first to have investigated the Case 2 in distributed optimization.

B. Game Under Partial-Decision Information

Consider a noncooperative $N$-player game with unconstrained action sets, where each player can be viewed as an agent and a Nash equilibrium is assumed to exist for the game. In this problem, each player $i \in [N]$ possesses its own cost (or payoff) function $J_i(x_i, x_{-i})$, which is differentiable, where $x_i$ is the decision/action vector of player $i$ and $x_{-i}$ denotes the decision vectors of all other players, i.e., $x_{-i} := \text{col}(x_{1,k}, \ldots, x_{i-1,k}, x_{i+1,k}, \ldots, x_{N,k})$. Note that player $i$ cannot access other players’ decision vectors, i.e., the considered game here is under partial-decision information, which is more practical than the case where each player has full access to all other players’ decisions in most of existing works. For this problem, at time step $k \geq 0$, each player $i \in [N]$ will choose its own decision vector $x_{i,k} \in \mathbb{R}^{n_i}$ and a cost $J_i(x_{i,k}, x_{-i,k})$ will be incurred for player $i$ after all players make their decisions. Then, the objective is for each player to minimize its own cost function, i.e., all players desire to achieve a Nash equilibrium $x^* = \text{col}(x_{1}^*, \ldots, x_{N}^*) \in \mathbb{R}^n$ with $n := \sum_{i=1}^{N} n_i$, which is defined as: For all $i \in [N]$

$$J_i(x_i^*, x_{-i}^*) \leq J_i(x_i, x_{-i}^*), \quad \forall x_i \in \mathbb{R}^{n_i}. \quad \text{(36)}$$

To proceed, let $\nabla J_i(x_i, x_{-i})$ denote $\nabla J_i(x_i, x_{-i})$ for simplicity. It is then easy to see that a Nash equilibrium $x^*$ satisfies $\nabla J_i(x_i^*, x_{-i}^*) = 0$ for all $i \in [N]$. Consequently, the Nash equilibrium seeking can be equivalently recast as to find fixed points of an operator $F$, defined by

$$F := Id - rU, \quad \text{(37)}$$

$$U := \text{col}(\nabla J_1, \ldots, \nabla N J_N) \quad \text{(38)}$$

where $r > 0$ is any constant. By defining $F_i := Id - r \nabla J_i$ for $i \in [N]$ with $Id_i : x \mapsto x_i$ for $x = \text{col}(x_1, \ldots, x_N) \in \mathbb{R}^n$, one can obtain that $F$ is block separable, i.e., $F = \text{col}(F_1, \ldots, F_N)$, which is consistent with problem (6). As a result, the linear convergence to a Nash equilibrium of the game can be assured by Theorem 2 under Assumptions 1–3 with $F$ being quasi-nonexpansive.

Note that Assumptions 1–3 are relatively mild, some of which are less conservative than those employed in the literature, as remarked below.

1) Assumption 2 is in fact equivalent to $\nabla J_i$ being Lipschitz for all $i \in [N]$, i.e., $\|\nabla J_i(x) - \nabla J_i(y)\| \leq q_i \|x - y\|$ for some $q_i > 0$ and for all $x, y \in \mathbb{R}^n$, which has been frequently employed in the literature, see e.g., [33], [47], [48], [49]. Then it can be readily obtained that $U$ is $q$-Lipschitz, where $q := \sqrt{\sum_{i=1}^{N} q_i^2}$.

2) The linear regularity is strictly weaker than the strong monotonicity, which has been widely imposed for deriving the linear convergence [33], [47], [48], [49], i.e., $(U(x) - U(y))^T (x - z) \geq \mu \|x - z\|^2$ for some $\mu > 0$ and for all $x, z \in \mathbb{R}^n$. To see this, it is obvious that strong monotonicity is strictly stronger than quasi-strong monotonicity, i.e., $(U(x) - U(y))^T (x - y) \geq \mu \|x - y\|^2$ for all $x \in \mathbb{R}^n$ and $y \in Fix(F)$. Meanwhile, quasi-strong monotonicity can imply the linear regularity of $F$, since it holds that $\|F(x) - z\| = r\|U(x)\| = r\|U(x) - U(F_{Fix(F)}(x))\| \geq r\mu \|x - F_{Fix(F)}(x)\| = r\mu d_{Fix(F)}(x)$ for all $x \in \mathbb{R}^n$, i.e., $d_{Fix(F)}(x) \leq \|F_{0}(x) - x\|/(r\mu)$, where $U(F_{Fix(F)}(x)) = 0$ and the quasi-strong monotonicity have been utilized. It should be also noted that the game can have a closed convex set of Nash equilibria (not necessary to be unique) under linear regularity.

3) The quasi-nonexpansiveness of $F$ is a weak assumption. For example, if the aforementioned quasi-strong monotonicity holds, then it holds that for all $x \in \mathbb{R}^n$ and $y \in Fix(F)$

$$\|F(x) - y\|^2 = \|x - y - r(U(x) - U(y))\|^2$$

$$= \|x - y\|^2 - 2r(x - y)^T (U(x) - U(y)) + r^2 \|U(x) - U(y)\|^2$$

$$\leq (1 - 2r + q^2 r^2) \|x - y\|^2 \quad \text{(39)}$$
where the quasi-strong monotonicity and $q$-Lipschitz continuity of $U$ are used in the inequality. In view of (39), it is easy to see that $F$ is even contractive, which is stronger than quasi-nonexpansive, if $r \in (0, \frac{2\mu}{q})$.

4) Assumption 1 requires the strong connectivity for directed graphs, which are not necessarily balanced. In contrast, balanced undirected/directed graphs are exploited in [33], [47], [48], and [49]. We note that time-varying graphs are considered in [33], but the graphs still need to be balanced, and in this case, it is interesting for us to extend the results of this article to time-varying communication graphs.

VI. NUMERICAL EXAMPLES

This section is to provide two numerical examples to corroborate the proposed algorithms.

**Example 1:** Consider a distributed optimization problem as discussed in Case 1 of Section V-A, where $f_i(x) = h_i(E x) + b_i^\top x$, and $h_i(z)$ is a strongly convex function with Lipschitz continuous gradient. It is easy to see that this problem is equivalent to finding a fixed point of the operator $F := I - \xi \nabla f$ for $\xi \in (0, 2/L)$ (see Section V-A), which is in the form (5) with $F_i := I - \xi \nabla f_i$ for $i \in [N]$.

It should be noted that $f_i$ will be strongly convex when $E$ has full column rank, and $f_i$ will be convex but not strongly convex if $E$ does not have full column rank ([30], Section III), which is frequently encountered in practical applications, such as the $L_1$-loss linear support vector machine (SVM) in machine learning [50]. Denote by $X^*$ the nonempty set of optimizers of this problem. Although $f_i$ is not strongly convex when $E$ does not have full column rank, it has been shown in [50], Theorem 18] that this problem admits a global error bound, i.e., $d_{X^*}(x) \leq \tau \|\nabla f(x)\|$, $\forall x \in \mathbb{R}^n$ for some constant $\tau \geq 0$, which further leads to $d_{F(x)}(x) \leq \frac{\tau}{\xi} \|x - F(x)\|$ for all $x \in \mathbb{R}^n$, i.e., satisfying the linear regularity condition.

In the simulation, let $N = 100$, $n = 5$, $h_i(E x) = |E x - p_i|^2$, $E = (1, 1, 1, 1, 1) \in \mathbb{R}^{1 \times 5}$, $p_i = i$, and $b_i = \text{col}(i, i, i, i, i)/5$ for all $i \in [N]$. Then, it is easy to verify that the gradient constant of $f$ is $L = 10$, thus holding $\xi \in (0, 0.2)$. Setting $\alpha = 0.004$ and $\xi = 0.15$, running the DOT algorithm (9) gives rise to the simulation results in Fig. 1, indicating that all $x_{i,k}$’s converge linearly to the optimal set $X^* := \{z = 5$ for $i,j \in [N]$ by DOP. [21].

**Example 2:** Consider the class of games as discussed in Section V-B with $N = 50$ players. To be specific, each player $i$ has its decision vector in $\mathbb{R}^2$ with its cost function given as $J_i(x_i, x_{-i}) = h_i(E x_i) + l_i^\top x_i$, where $h_i(z) = r_i z^2 + s_i z$ is strongly convex for $z \in \mathbb{R}$ with $r_i > 0$ and $s_i \in \mathbb{R}$, and $l_i(x_{-i}) = \sum_{j \neq i} c_{ij} x_j$ with $c_{ij} \in \mathbb{R}$ and $c_{ij} \neq 0$. Note that $J_i$ is not strongly convex in $x_i$ if $E$ is not of full column rank, as discussed in Example 1.

In this example, let $E = (1, 1)$, which does not have full column rank. Thus, $J_i$ is not strongly convex in $x_i$. However, it is easy to verify that $F_i = I - r \nabla J_i$ is Lipschitz, i.e., satisfying Assumption 2. Meanwhile, it can be claimed that the global operator $F$, as defined in (37), is linearly regular for small $r > 0$, but may be not quasi-nonexpansive, whose proofs are postponed to the appendix for the clarity of presentation.

Set $\alpha = 0.001, r = 0.1, c_{ij} = 0$ when $i < j$, and $c_{ij} = c_{ij} I$ for $c_{ij} \in \mathbb{R}$ when $i > j$. By randomly choosing $r_i$, $s_i$, and $c_{ij}$ for $i,j \in [N]$ with a randomly generated strongly connected
communication graph, performing the developed DOP with each component of initial conditions randomly in [10,20] gives the simulation results in Figs. 3 and 4. In Fig. 3, the trajectories of $\|x_{d,k}^i - x_{d,k}^j\|$ are plotted for all players, showing that all $x_{d,k}^i$’s achieve consensus at a linear rate. Meanwhile, the gradient of each agent $i$ is given in Fig. 4, indicating that all agents’ gradients converge linearly to zero. From the two figures, it is interesting to observe that all players’ estimates $x_{d,k}^i$’s converge to the equilibrium set at a linear rate even though $F$ may be not quasi-nonexpansive. In summary, the simulation results support the theoretical result in Theorem 2, simultaneously inspiring that the quasi-nonexpansiveness of $F$ may be relaxed in future.

VII. CONCLUSION

This article has investigated the fixed-point seeking problem for a quasi-nonexpansive global operator over a fixed, unbalanced, and directed communication graph, for which two scenarios have been considered, i.e., the global operator is respectively sum separable and block separable under the regularity condition. For the first case, the global operator consists of a sum of local operators, which are assumed to be Lipschitz. To solve this case, a distributed algorithm, DOT, has been proposed and shown to be convergent to a fixed point of the global operator at a linear rate. For the second case, a distributed algorithm, DOP, has been developed, showing to be linearly convergent to a fixed point of the global operator. Meanwhile, two applications have been presented in detail, i.e., distributed optimization and multiplayer game under partial-decision information. In future, it is interesting to study asynchronous algorithms, nonidentical step sizes for agents, time-varying communication graphs, and possible relaxation of the quasi-nonexpansiveness of the global operator.

APPENDIX

A. Proof of Lemma 2

To prove (21), it can be obtained that

$$\|A z - A_\infty z\|_\pi = \|(A - A_\infty)(z - A_\infty z)\|_\pi \leq \|A - A_\infty\|_\pi \|z - A_\infty z\|_\pi$$

(40)

where the equality has used the fact $AA_\infty = A_\infty A_\infty = A_\infty$. Consider the term $\|A - A_\infty\|_\pi$ in (40). To do so, by definition (13), one has that for any $x \in \mathbb{R}^N$ and $y \in \mathcal{H}$

$$\|x \otimes y\|_\pi = \sqrt{\sum_{i=1}^N \pi_i x_i^2 y_i^2} = \|x\| \sqrt{\sum_{i=1}^N \pi_i x_i^2}$$

(41)

which, together with the norm’s definition, leads to

$$\|A - A_\infty\|_\pi = \sup_{\|x \otimes y\|_\pi \neq 0} \frac{\|(A - A_\infty)(x \otimes y)\|_\pi}{\|x \otimes y\|_\pi}$$

$$= \sup_{\|x \otimes y\|_\pi \neq 0} \frac{\|(A - A_\infty)x \|_\pi y\|_\pi}{\|x\| \|y\|_\pi}$$

$$= \sup_{\|x \otimes y\|_\pi \neq 0} \frac{\|(A - A_\infty)x\|_\pi}{\|x\|}$$

(42)

Note that $\|A - A_\infty\|_\pi = \rho_1$ by Lemma 1. Consequently, putting together (40)–(42) gives rise to (21). By noting that $BB_\infty = B_\infty B_\infty = B_\infty$, similar arguments can be applied to obtain (22) and (23). This ends the proof.

B. Proof of Theorem 1

Let us first establish upper bounds on $\|x_{k+1} - A_\infty x_{k+1}\|_\pi$, $\|x_{k+1} - x_k\|_\pi$, $\|y_{k+1} - B_\infty y_{k+1}\|_\pi$, and $\|A_\infty x_{k+1} - 1_N \otimes x_{k+1}\|_\pi$. Note that $\bar{x}_k := \sum_{i=1}^N \pi_i x_{i,k}$ and $x_{d,k}^i := P_{F(x)}(\bar{x}_k)$ for all $k \geq 0$. For clarity, the proof is given in five steps.

**Step 1. To bound $\|x_{k+1} - A_\infty x_{k+1}\|_\pi$.** By noting $A_\infty A = A_\infty$, invoking (10) yields that

$$\|x_{k+1} - A_\infty x_{k+1}\|_\pi = \|(1 - \alpha)A x_k + \frac{\alpha}{N}(D_{\nu^{-1}} \otimes I) y_k - (1 - \alpha)A_\infty A x_k\|_\pi$$

$$\leq (1 - \alpha)\|A x_k - A_\infty x_k\|_\pi + \frac{\alpha}{N}\|(I_N \otimes I - A_\infty)(D_{\nu^{-1}} \otimes I) y_k\|_\pi.$$ (43)

Consider the last term in (43). One can obtain that

$$\|(I_N \otimes I - A_\infty)(D_{\nu^{-1}} \otimes I) y_k\|_\pi = \|(I_N \otimes I - A_\infty)(D_{\nu^{-1}} \otimes I) y_k - 1_N \otimes \bar{y}_k\|_\pi$$

$$= \|(I_N \otimes I - A_\infty)(D_{\nu^{-1}} \otimes I) y_k - \nu \otimes \bar{y}_k\|_\pi,$$

$$\leq \|I_N \otimes I - A_\infty\|_\pi \|D_{\nu^{-1}} \otimes I\|_\pi \|y_k - B_\infty y_k\|_\pi$$

$$\leq c_k \|D_{\nu^{-1}}\|_\pi \|y_k - B_\infty y_k\|_\nu.$$ (44)

where $\bar{y}_k$ is defined in Lemma 3, the first inequality has applied the fact that $\nu \otimes \bar{y}_k = B_\infty \bar{y}_k$, and the last inequality has employed (23), $\|D_{\nu^{-1}} \otimes I\|_\pi = \|D_{\nu^{-1}}\|_\pi$ (using the same argument as that in Lemma 2), and (16).
In view of (15) and (21), inserting (44) in (43) results in
\[
\|x_{k+1} - A_\infty x_{k+1}\|_\pi \leq (1 - \alpha) \rho_1 \|x_k - A_\infty x_k\|_\pi + \frac{\alpha c_2 \|D_{\nu}^{-1}\|}{N}\|y_k - B_{\infty} y_k\|_\nu.
\]
(45)

**Step 2. To bound \(\|x_{k+1} - x_k\|\):** Invoking (10) yields that
\[
\|x_{k+1} - x_k\| = \|A x_k - x + \alpha \left( \frac{D_{\nu}^{-1} \otimes I_d}{N} y_k - A x_k \right)\|
\leq \|A - I\| \|x_k - A_\infty x_k\| + \alpha \frac{\|D_{\nu}^{-1}\|}{N}\|y_k - B_{\infty} y_k\|
+ \alpha \frac{\|D_{\nu}^{-1} \otimes I_d\|}{N} B_{\infty} y_k - A x_k\|
\]
where the inequality has leveraged the triangle inequality and the facts that \((A - I_N \otimes I_d)(x_k - A_\infty x_k) = A x_k - x_k\) and \(\|A - I_N \otimes I_d\| = \|A - I\|\) (using the same argument as that in Lemma 2).

Consider the last term in (46). By using \(B_{\infty} = \nu 1_N \otimes I_d\), one has that
\[
\|\frac{D_{\nu}^{-1} \otimes I_d}{N} B_{\infty} y_k - A x_k\| = \|1_N \otimes \frac{y_k}{N} - A x_k\|
\leq \|1_N \otimes \left( \frac{y_k}{N} - \sum_{i=1}^{N} F_i(\bar{x}_k) \right)\|
+ \|1_N \otimes \left( \sum_{i=1}^{N} F_i(\bar{x}_k) - x_k\right)\|
+ \|1_N \otimes x_k - A_\infty x_k\| + \|A_\infty x_k - A x_k\|.
\]
(47)

For the first term in the last inequality in (47), invoking Lemma 3, one can obtain that
\[
\|1_N \otimes \left( \frac{y_k}{N} - \sum_{i=1}^{N} F_i(\bar{x}_k) \right)\|^2
= \|1_N \otimes \sum_{i=1}^{N} \left( F_i(x_{i,k}) - F_i(\bar{x}_k) \right)\|^2
\leq \sum_{i=1}^{N} \|F_i(x_{i,k}) - F_i(\bar{x}_k)\|^2
\leq \sum_{i=1}^{N} L_i^2 \|x_{i,k} - \bar{x}_k\|^2
\leq \bar{L}^2 \|x_k - A_\infty x_k\|^2
\]
(48)

where the first and second inequalities have employed \(\|\sum_{i=1}^{N} z_i\|^2 \leq N \sum_{i=1}^{N} \|z_i\|^2\) for any vectors \(z_i\)'s and Assumption 2, respectively. Similarly, it can be obtained that
\[
\|1_N \otimes \left( \sum_{i=1}^{N} F_i(\bar{x}_k) - x_k\right)\|^2 = N \|F(\bar{x}_k) - x_k\|^2
\leq N \|\bar{x}_k - x_k\|^2
\leq N \|x_{k+1} - x_k\|^2
\]
(49)

where \(x_k^* \in Fix(F)\) and the quasi-nonexpansiveness of \(F\) have been used in the inequality.

As a result, substituting (48) and (49) into (47) leads to
\[
\|x_{k+1} - x_k\| \leq \alpha (\rho_1 + \bar{L}) + \|A - I\| \|x_k - A_\infty x_k\| + \frac{\alpha c_2 \|D_{\nu}^{-1}\|}{N} \|y_k - B_{\infty} y_k\|_\nu
+ \frac{\|D_{\nu}^{-1} \otimes I_d\|}{N} B_{\infty} y_k - A x_k\|
\]
(50)

where (15) and (17) have been utilized in the last inequality.

Putting (50) in (46) leads to
\[
\|x_{k+1} - x_k\| \leq \alpha (\rho_1 + \bar{L}) + \|A - I\| \|x_k - A_\infty x_k\| + \frac{\alpha c_2 \|D_{\nu}^{-1}\|}{N} \|y_k - B_{\infty} y_k\|_\nu
+ \frac{\|D_{\nu}^{-1} \otimes I_d\|}{N} B_{\infty} y_k - A x_k\|
\]
(51)

**Step 3. To bound \(\|y_{k+1} - B_{\infty} y_k\|_\nu\):** Invoking (11) yields
\[
\|y_{k+1} - B_{\infty} y_{k+1}\|_\nu = \|B y_k - B_{\infty} B y_k + F(x_{k+1}) - B y_{k+1}\|_\nu
\leq \|B y_k - B_{\infty} y_k\|_\nu + \|F(x_{k+1}) - F(x_k)\|_\nu
+ \|B_{\infty} (F(x_{k+1}) - F(x_k))\|_\nu
\leq \rho_2 \|y_k - B_{\infty} y_k\|_\nu + \frac{c_2 (1 + \sqrt{N})}{c_3} \|F(x_{k+1}) - F(x_k)\|_\nu
\]
(52)

where \(B_{\infty} B = B_{\infty}\) has been used in the first inequality, and (15), (16), (22), and \(\|B_{\infty}\| = \|B_{\infty}\| \leq \sqrt{N}\|B_{\infty}\| \leq \sqrt{N}\) have been exploited in the last inequality.

On the other hand, it can be obtained that
\[
\|F(x_{k+1}) - F(x_k)\|^2 = \sum_{i=1}^{N} \|F_i(x_{i,k+1}) - F_i(x_{i,k})\|^2
\leq \sum_{i=1}^{N} L_i^2 \|x_{i,k+1} - x_{i,k}\|^2
\leq \bar{L}^2 \|x_{k+1} - x_k\|^2
\]
(53)

where Assumption 2 has been applied to obtain the first inequality in (53).

Now substituting (51) and (53) into (52) leads to
\[
\|y_{k+1} - B_{\infty} y_{k+1}\|_\nu \leq \rho_2 \|y_k - B_{\infty} y_k\|_\nu + \frac{c_2 \bar{L} (\sqrt{N} + 1)}{c_3} \|x_{k+1} - x_k\|
\leq \left( \rho_2 + \frac{\alpha c_2 \bar{L} (\sqrt{N} + 1) \|D_{\nu}^{-1}\|}{N c_1 c_3} \right) \|y_k - B_{\infty} y_k\|_\nu
\]

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
\[
\begin{align*}
+ \frac{c_2 L (\sqrt{N} + 1)}{c_4} \left( A_\infty - x_k \right) & \left( x_k - A_\infty x_k \right)_\pi \\
+ \frac{2c_2 L (\sqrt{N} + 1)}{c_3} \left[ A_\infty x_k - 1_N \otimes x_k^* \right] & \leq \frac{\theta_3}{c_1} \left( x_k - A_\infty x_k \right) \left( x_k - A_\infty x_k \right)_\pi
\end{align*}
\]

where \( \theta_3 = \rho_1 + L. \)

Step 4. To bound \( \left\| A_\infty x_{k+1} - 1_N \otimes x_{k+1}^* \right\| \): By defining \( x_k^* := P_{Fix}(F_\alpha(x_k)) \) and noting that \( A_\infty x_{k+1} = 1_N \otimes x_{k+1}^* \) and \( A_\infty \mathbf{A} = A_\infty \), invoking (10) gives rise to

\[
\begin{align*}
\left\| A_\infty x_{k+1} - 1_N \otimes x_{k+1}^* \right\| & \leq \left\| A_\infty x_k + \alpha \left[ A_\infty \left( D_{\nu}^{-1} \otimes \text{Id} \right) y_k - A_\infty x_k \right] - 1_N \otimes x_k^* \right\| \\
& = \left\| 1_N \otimes x_k + \alpha \left[ A_\infty \left( 1_N \otimes y_k \right) - 1_N \otimes x_k \right] - 1_N \otimes x_k^* \right\| \\
& + \alpha \left\| A_\infty \left( D_{\nu}^{-1} \otimes \text{Id} \right) (y_k - B_\infty y_k) \right\| \\
& \leq \left\| 1_N \otimes x_k + \alpha \left[ A_\infty \left( 1_N \otimes y_k \right) - 1_N \otimes x_k \right] - 1_N \otimes x_k^* \right\| \\
& + \frac{\alpha c_4 \left\| D_{\nu}^{-1} \right\|}{\sqrt{N} c_1} \left\| y_k - B_\infty y_k \right\|_\nu
\end{align*}
\]

where the last inequality has utilized (15), (16), and the fact that \( \left\| A_\infty \right\| = \left\| A_\infty \right\| \leq \sqrt{\left\| A_\infty \right\| \left\| A_\infty \right\|} \leq \sqrt{N}. \)

On the other hand, by \( A_\infty = 1_N \otimes \pi^* \) and Lemma 3, one can obtain that

\[
\begin{align*}
\left\| 1_N \otimes x_k + \alpha \left[ A_\infty \left( 1_N \otimes y_k \right) - 1_N \otimes x_k \right] - 1_N \otimes x_k^* \right\| & = \left\| 1_N \otimes x_k + \alpha \left[ 1_N \otimes \sum_{i=1}^N F_i(x_{i,k}) - 1_N \otimes x_k \right] - 1_N \otimes x_k^* \right\| \\
& = \left\| 1_N \otimes F_\alpha(x_k) - 1_N \otimes x_k^* \right\| \\
& + \alpha \left\| 1_N \otimes \sum_{i=1}^N \left( F_i(x_{i,k}) - F_i(x_k) \right) \right\| \\
& \leq \left\| 1_N \otimes \left( F_\alpha(x_k) - x_k^* \right) \right\| \\
& + \alpha \left\| 1_N \otimes \sum_{i=1}^N \left( F_i(x_{i,k}) - F_i(x_k) \right) \right\|
\end{align*}
\]

where \( F_\alpha \) is defined in (7).

For the term \( \left\| 1_N \otimes \left[ F_\alpha(x_k) - x_k^* \right] \right\| \) in (56), invoking \( Fix(F_\alpha) = Fix(F) \) and Lemma 4 implies that

\[
\left\| 1_N \otimes \left[ F_\alpha(x_k) - x_k^* \right] \right\|^2 = N \left\| F_\alpha(x_k) - x_k^* \right\|^2 \\
\leq N \rho_3^2 \left\| x_k - P_{Fix} \left( x_k \right) \right\|^2 \\
= N \rho_3^2 \left\| x_k - x_k^* \right\|^2 \\
= \rho_3^2 \left\| 1_N \otimes \left( x_k - x_k^* \right) \right\|^2.
\]

Now, putting together (48) and (55)–(57) results in

\[
\begin{align*}
\left\| A_\infty x_{k+1} - 1_N \otimes x_{k+1}^* \right\| & \leq \rho_3 \left\| A_\infty x_k - 1_N \otimes x_k^* \right\| + \frac{\alpha L}{c_1} \left\| x_k - A_\infty x_k \right\| \left\| x_k - A_\infty x_k \right\| \\
& + \frac{\alpha c_4 \left\| D_{\nu}^{-1} \right\|}{\sqrt{N} c_1} \left\| y_k - B_\infty y_k \right\|_\nu.
\end{align*}
\]

Step 5. To show the linear convergence: Define \( z_k := \text{col}(x_k - A_\infty x_k, y_k - B_\infty y_k, A_\infty x_k - 1_N \otimes x_k^*) \). Combining (45), (54), and (58) with \( \alpha \in (0, 1) \) yields that

\[
\begin{align*}
z_{k+1} & \leq M(\alpha) z_k
\end{align*}
\]

where \( M(\alpha) \) is defined in (27).

It is easy to see that \( z_k \) will converge to the origin at a linear rate if \( \rho(M(\alpha)) < 1 \). To ensure \( \rho(M(\alpha)) < 1 \), it is straightforward to observe that when \( \alpha = 0 \) is a simple eigenvalue of \( M(0) \) with corresponding left and right eigenvectors both being \( \text{col}(0, 0, 1) \). Then invoking [45, Lemma 5] gives rise to

\[
\left. \frac{dx(\alpha)}{d\alpha} \right|_{\alpha=0} = -\frac{\delta}{2\alpha^2} < 0
\]

indicating that the simple eigenvalue 1 of \( M(0) \) will decrease when increasing the value of \( \alpha \). Thus, by the continuity of \( \rho(M(\alpha)) \) with respect to \( \alpha \), there must exist a constant \( \alpha_c > 0 \) such that \( \rho(M(\alpha)) < 1 \) for all \( \alpha \in (0, \alpha_c) \). To find \( \alpha_c \), one can see that the graph associated with \( M(\alpha) \) consisting of three agents is strongly connected, which, in conjunction with [14, Theorem C.3], leads to \( M(\alpha) \) being irreducible. Together with the fact that \( M \) has at least one nonzero diagonal entry, it can be obtained that \( M(\alpha) \) is primitive, which results in that \( \rho(M(\alpha)) \) is a simple eigenvalue of \( M(\alpha) \) and all other eigenvalues have absolute values of less than \( \rho(M(\alpha)) \). Moreover, it can be ensured that \( \rho(M(\alpha)) = 1 \) when increasing \( \alpha \) from 0 to some point, and thereby the value of \( \alpha_c \) can be calculated by letting \( \det(I - M(\alpha)) = 0 \). Therefore, when \( \alpha \in \text{min}\{1 - \delta, \alpha_c\} \), it holds that \( \rho(M(\alpha)) < 1 \).

For any \( \epsilon > 0 \), by [51, Lemma 5.6.10], there exists a norm \( \left\| \cdot \right\| \) in \( \mathbb{R}^3 \) such that \( \left\| M(\alpha) \right\| \leq \rho(M(\alpha)) + \epsilon \). Let us choose a small enough \( \epsilon \) such that \( \rho(M(\alpha)) + \epsilon < 1 \). Then, invoking (59) gives rise to

\[
\left\| z_k \right\| \leq \left\| M(\alpha) \right\|^{k} \left\| z_0 \right\| \leq \gamma^{k} \left\| z_0 \right\|, \quad \forall k \geq 0
\]

where \( \gamma := \rho(M(\alpha)) + \epsilon \in (0, 1) \). Since all norms in finite dimensional spaces are equivalent, there exist two constants \( \varsigma_1, \varsigma_2 > 0 \) such that \( \varsigma_1 \| \cdot \| \leq \| \cdot \| \leq \varsigma_2 \| \cdot \| \). Together with (60), it can be obtained that

\[
\left\| z_k \right\| \leq \varsigma_2 \left\| z_k \right\| \leq \varsigma_2 \gamma^{k} \left\| z_0 \right\|, \quad \forall k \geq 0
\]

which, combining with the fact that \( \| x_k - A_\infty x_k \|_\pi \leq \| z_k \| \) and \( \| A_\infty x_k - 1_N \otimes x_k^* \| \leq \| z_k \| \), yields

\[
\| x_k - A_\infty x_k \|_\pi \leq \varsigma_2 \gamma^{k} \left\| z_0 \right\|, \quad \forall k \geq 0.
\]

Consequently, for all \( k \geq 0 \), one has that

\[
\| x_k - 1_N \otimes x_k^* \| \leq \| x_k - A_\infty x_k \| + \| A_\infty x_k - 1_N \otimes x_k^* \|
\]
\[
\leq \frac{1}{c_1} \|x_k - A_{\infty} x_k\|_\pi + \|A_{\infty} x_k - 1_N \otimes x_k^i\|
\leq 2q(1 + c_1)^q_k \|z_0\|
\]
(63)
where (15) has been employed in the second inequality. This completes the proof.

C. Proof of Theorem 2

Let us bound \(\|x_{k+1} - A_{\infty} x_{k+1}\|_\pi\) and \(d_{Fix}(F) (\tilde{x}_{k+1})\) in the following. For clarity, the proof is given in three steps.

Step 1. To bound \(\|x_{k+1} - A_{\infty} x_{k+1}\|_\pi\). In view of (29) and \(A_{\infty} A = A_{\infty}\), one has that

\[
\|x_{k+1} - A_{\infty} x_{k+1}\|_\pi = \|Ax_k + \alpha \tilde{F} - A_{\infty} Ax_k - \alpha A_{\infty} \tilde{F}\|_\pi
\leq \|Ax_k - A_{\infty} x_k\|_\pi + \alpha \|\tilde{F} - 1_N \tilde{F}\|_\pi
\leq \rho_1 \|x_k - A_{\infty} x_k\|_\pi + \alpha c_2 \|\tilde{F} - 1_N \tilde{F}\|
\]
(64)
where (15) and (31) have been utilized in the last inequality.

For the last term in (64), it is easy to verify that

\[
\tilde{F} - 1_N \tilde{F} = \bigg(\frac{1}{\pi_i} - 1\bigg) e_{i,k} e_{2,k} \cdots e_{N,k}
\]
with \(e_{i,k} := F_i(x_k^i) - \tilde{x}_{i,k}\), and thus it can be obtained that

\[
\|\tilde{F} - 1_N \tilde{F}\|_\pi^2 = \sum_{i=1}^N \left(\frac{1}{\pi_i} - 1\right)^2 \|e_{i,k}\|^2 + \sum_{j \neq i} \|e_{j,k}\|^2
\]
\begin{align}
\leq 2c_1 \sum_{i=1}^N \|e_{i,k}\|^2 \\
= 2c_1 \sum_{i=1}^N \left(\|F_i(x_k^i) - F_i(\tilde{x}_k) + F_i(\tilde{x}_k) - x_i^i + x_i^i - \tilde{x}_{i,k}\right.
+ \tilde{x}_{i,k} - \tilde{x}_{i,k}\|^2
\leq 4c_1 \sum_{i=1}^N \left(\|F_i(x_k^i) - F_i(\tilde{x}_k)\|^2 + \|F_i(\tilde{x}_k) - x_i^i\|^2
+ \|\tilde{x}_{i,k} - x_i^i\|^2 + \|\tilde{x}_{i,k} - \tilde{x}_{i,k}\|^2\right)
\end{align}
(65)
where \(\omega := N - 1 + (1 - \pi)^2/\pi^2\), \(x^* = (x_1^* , \ldots , x_N^*)\) denotes any fixed point of \(F\), \(\pi_i \geq \pi\) has been used in the first inequality and \(\|\sum_{i=1}^n z_i\|^2 \leq n \sum_{i=1}^n \|z_i\|^2\) for any vectors \(z_i\)'s in the last inequality. Note that \(\tilde{x}_{i,k}\) is the \(i\)th block-coordinate of \(\tilde{x}_k\) defined in (30).

To proceed, one can obtain that

\[
\sum_{i=1}^N \|\tilde{x}_{i,k} - \tilde{x}_{i,k}\|^2 = \sum_{i=1}^N \sum_{j=1}^N a_{ij} \|x_{i,k}^j - \tilde{x}_{i,k}\|^2
\leq \sum_{i=1}^N \sum_{j=1}^N a_{ij} \|x_{i,k}^j - \tilde{x}_{i,k}\|^2
\leq \sum_{i=1}^N \sum_{j=1}^N \|x_{i,k}^j - \tilde{x}_{i,k}\|^2
= \|x_k - 1_N \tilde{x}_k\|^2
\]
(66)
where \(\sum_{j=1}^N a_{ij} = 1\) has been exploited for the first equality, the convexity of norm \(\cdot\|\cdot\|^2\) for the first inequality, and \(a_{ij} \leq 1\) for the second inequality.

Now, invoking Assumptions 2 and 3.1. (65) and (66) yields

\[
\|\tilde{F} - 1_N \tilde{F}\|^2 \leq 4c_1 (L + 1) \|x_k - 1_N \tilde{x}_k\|^2
+ 8c_1 \|\tilde{x}_k - x^*\|^2
\]
which, together with \(1_N \tilde{x}_k = A_{\infty} x_k\), implies

\[
\|\tilde{F} - 1_N \tilde{F}\|^2 \leq 2\sqrt{c_1(L + 1)} \|x_k - A_{\infty} x_k\|
+ 2\sqrt{2c_1 \|\tilde{x}_k - x^*\|}
\]
(68)

At this step, by choosing \(x^* = P_{Fix}(F)(\tilde{x}_k)\), substituting (68) into (64) leads to

\[
\|x_{k+1} - A_{\infty} x_{k+1}\|_\pi \leq (\rho_1 + \alpha \theta_0) \|x_k - A_{\infty} x_k\|
+ 2\alpha c_2 \sqrt{2c_1 d_{Fix}(F)(\tilde{x}_k)}
\]
(69)
where \(\theta_0 := 2c_2 \sqrt{N(L + 1)/\pi c_1}\).

Step 2. To bound \(d_{Fix}(F)(\tilde{x}_{k+1})\): One can observe that

\[
\tilde{F} = (e_{1,k} , \ldots , e_{N,k})
= (F_1(\tilde{x}_k) - \tilde{x}_{1,k} , \ldots , F_N(\tilde{x}_k) - \tilde{x}_{N,k})
+ h_{1,k} + h_{2,k}
\]
(70)
where \(e_{i,k} := F_i(x_k^i) - \tilde{x}_{i,k}\) and

\[
h_{1,k} := (F_1(x_k^1) - F_1(\tilde{x}_k) , \ldots , F_N(x_k^N) - F_N(\tilde{x}_k)),
\]
\[
h_{2,k} := (\tilde{x}_{1,k} - \tilde{x}_{1,k} , \ldots \tilde{x}_{N,k} - \tilde{x}_{N,k})
\]
Meanwhile, invoking Assumption 2 yields that

\[
\|h_{1,k}\|^2 = \sum_{i=1}^N \|F_i(x_k^i) - F_i(\tilde{x}_k)\|^2
\leq \sum_{i=1}^N L_i^2 \|x_k^i - \tilde{x}_k\|^2
\leq \tilde{L} \|x_k - 1_N \tilde{x}_k\|^2
\]
(71)
and by (66) and \(\pi_i \in (0, 1)\), it has that

\[
\|h_{2,k}\|^2 = \sum_{i=1}^N \|\tilde{x}_{i,k} - \tilde{x}_{i,k}\|^2 \leq \|x_k - 1_N \tilde{x}_k\|^2.
\]
(72)

Now, in view of (30), (70), (71), (72), and \(1_N \tilde{x}_k = A_{\infty} x_k\), it has that for \(y^* = P_{Fix}(F)(\alpha(\tilde{x}_k)) \in Fix(F)\)

\[
\|\tilde{x}_{k+1} - y^*\| = \|\tilde{x}_k + \alpha \tilde{F} - y^*\|
\leq \|F_\alpha(\tilde{x}_k) - y^* + \alpha(h_{1,k} + h_{2,k})\|
\leq \|F_\alpha(\tilde{x}_k) - y^*\| + \alpha(h_{1,k} + h_{2,k})
\leq d_{Fix}(F)(F_\alpha(\tilde{x}_k)) + \frac{\alpha(L + 1)}{c_1} \|x_k - A_{\infty} x_k\|_\pi
\]
(73)
where (15) has been employed in the last inequality.
To analyze the term $d_{F(x)}(F(x_i)(x_i - x_{i-1}))$ in (73), invoking Lemma 4 and (73) yields that
\[
\|x_{k+1} - y^*\| \\
\leq \rho_3 d_{F(x)}(\hat{x}) + \frac{\alpha(\hat{L} + 1)}{c_1} \|x_k - A_{x_k}x_k\|. 
\]
Combining (73) with $d_{F(x)}(\hat{x}) + \|x_{k+1} - y^*\|$ can yield that
\[
d_{F(x)}(\hat{x}) + \|x_{k+1} - y^*\| \leq \frac{\alpha(\hat{L} + 1)}{c_1} \|x_k - A_{x_k}x_k\|. 
\]
(74)

Step 3. To show the linear convergence: By setting $z_k := \text{col}(x_k - A_{x_k}x_k, d_{F(x)}(\hat{x}))$ for $k \geq 0$, invoking (69) and (75) results in
\[
z_{k+1} \leq \Theta(\alpha)z_k 
\]
where $\Theta(\alpha)$ is defined in (33). Note that $A_{x_k}x_k = 1_N \cdot x_k$. At this step, Theorem 2 can be proved by following the similar analysis to that after (59) for proving Theorem 1.

A4 Proofs of F’s Properties in Example 2

1) The nonexpansiveness of $F$: Simple calculation yields that $F_i = I_{d_i} - r_i \nabla_i J_i = (I_{d_i} - 2r_i E^T E) x_i - s_i E^T - r_i (x_{i-1})$, and it can be then concluded that
\[
F(x) = Qx + rR 
\]
where $F$ is defined in (37), $R := \text{col}(s_1 E^T, \ldots, s_N E^T), Q := I_{2N} + rQ_1.$ and
\[
Q_1 := \begin{pmatrix}
-2r_1 E^T & -c_{12} & \cdots & -c_{1N} \\
-c_{21} & -2r_2 E^T & \cdots & -c_{2N} \\
\vdots & \vdots & \ddots & \vdots \\
-c_{N1} & -c_{N2} & \cdots & -2r_N E^T
\end{pmatrix}. 
\]
(78)

Therefore, one has that $\|F(x) - F(y)\| \leq \|Q\| \|x - y\|$ for any $x, y \in \mathbb{R}^{2N}$, thus implying that the nonexpansiveness (equivalent to quasi-nonexpansiveness since $F$ is linear) of $F$ closely depends upon $\|Q\|$. In the sequel, we shall that $\|Q\|$ may be greater than 1, thus indicating that $F$ may not quasi-nonexpansive. To do so, consider a specific case with $N = 2$, $r_1 = r_2$, and $c_{i,j} = c_{j,i}$ for some $c > 0$, where $I_{2}$ denotes the identity matrix of dimension 2. Note that the eigenvalues of $E^T E$ are 2 and 0 for $E = (1, 1)$. Then, a simple algebraic calculation yields that $Q_1$ has a positive eigenvalue $c$, and thereby, $Q$ has an eigenvalue $1 + rc$ that is greater than 1. Therefore, one can conclude that $\|Q\| \geq \rho(Q) \geq 1 + rc > 1$. This ends the proof.

2) The linear regularity of $F$: Consider the mapping $x_i \mapsto \nabla_i J_i(x_i, x_{i-1})$ with any fixed $x_{i-1}$ by each agent $i \in [N]$, which can be viewed as an operator with respect to $x_i$. It is easy to verify that $\nabla_i J_i(x_i, x_{i-1})$ is Lipschitz with the Lipschitz constant being independent of $x_{i-1}$, and $h_i(z) = r_i z^2 + s_i z$ is $r_i$-strongly convex. As a result, in light of [50, Theorem 18], there holds
\[
d_{X_i}(x_i) \leq \kappa_i \|\nabla_i J_i(x_i, x_{i-1})\|, \quad \forall i \in [N] 
\]
where $X_i := \{x_i : \nabla_i J_i(x_i, x_{i-1}) = 0\}$. $\kappa_i = \frac{1 + \rho_i}{\rho_i}$. and the constants $\rho_i, \theta$ are independent of $x_i, x_{i-1}$, depending only on $E, r_i, s_i$ and $E$, respectively. By defining $X^*_i := \{x_i : \nabla_i J_i(x_i, x_{i-1}) = 0\}$ and noticing that $x_{i-1}$ is any fixed vector, it is easy to see that $d_{X_i}(x_i) = d_{X_i}(x_{i-1})$. Therefore, (79) amounts to
\[
d_{X_i}(x_i) \leq \kappa_i \|\nabla_i J_i(x_i, x_{i-1})\|, \quad \forall i \in [N]. 
\]
(80)

Denote by $X^*$ the set of Nash equilibria, i.e., $X^* = \{x_i : \nabla_i J_i(x_i, x_{i-1}) = 0, \forall i \in [N]\}$. Obviously, it holds that $X^* = \bigcap_{i=1}^N X^*_i$. Note that $\nabla_i J_i(x_i, x_{i-1}) = 2r_i E^T E x_i + s_i E^T + l_i(x_{i-1})$ and thus, $X^*_i$ is a polyhedron. Thus, by Fact 5.8. (vii) in [29], it can be obtained that $\{X^*_i\}_{i=1}^N$ is linearly regular, i.e., $d_{x_i}^N X^*_i(x) \leq \mu_{\max_{i \in [N]}}(d_{X^*_i}(x))$ for some $\mu > 0$, which together with (80) and $\bigcap_{i=1}^N X^*_i = X^*$ leads to
\[
d_{X_i}(x_i) \leq \mu^2 \kappa^2_{\max} \|\nabla_i J_i(x_i, x_{i-1})\|^2 \\
\leq \mu^2 \kappa^2_{\max} \|U(x_i)\|^2 \\
= \frac{\mu^2 \kappa^2_{\max}}{r^2} \|x - F(x)\|^2 
\]
(81)

where $\kappa_{\max} := \max_{i \in [N]} \kappa_i$ and $F = I_2 - rU$ has been used for deriving the equality. Therefore, it can be concluded that $F$ is linearly regular by (81). This ends the proof.

ACKNOWLEDGMENT

The authors would like to thank the Editor, the Associate Editor, and the anonymous reviewers for their insightful suggestions. The authors also thank Mr. Yang Yu for his assistance in numerical simulations.

REFERENCES

[1] H. H. Bauschke and P. L. Combettes, Convex Analysis and Monotone Operator Theory in Hilbert Spaces, 2nd ed. New York, NY, USA: Springer, 2017.
[2] A. Cegielski, Iterative Methods for Fixed Point Problems in Hilbert Spaces, vol. 2057. Heidelberg, Germany: Springer, 2012.
[3] J. Liang, J. Fadili, and G. Peyré, “Convergence rates with inexact nonexpansive operators,” Math. Program., vol. 159, no. 1-2, pp. 403–434, 2016.
[4] J. M. Borwein, G. Li, and M. K. Tam, “Convergence rate analysis for averaged fixed point iterations in common fixed point problems,” SIAM J. Optim., vol. 27, no. 1, pp. 1–33, 2017.
[5] M. Bravo, R. Cominetti, and M. Pavez-Sigüe, “Rates of convergence for inexact Krasnosel’skiĭ-Mann iterations in Banach spaces,” Math. Program., no. 1–2, pp. 241–262, 2019.
[6] A. Themelis and P. Patrinos, “SuperMann: A superlinearly convergent algorithm for finding fixed points of nonexpansive operators,” IEEE Trans. Autom. Control, vol. 64, no. 12, pp. 4875–4890, Dec. 2019.
[7] W. R. Mann, “Mean value methods in iteration,” Proc. Amer. Math. Soc., vol. 4, no. 3, pp. 506–510, 1953.
[8] M. A. Krasnosel’ski˘, “Two comments on the method of successive approximations,” Uspekhi Matematicheskikh Nauk, vol. 10, pp. 123–127, 1955.
[9] S. Reich, “Weak convergence theorems for nonexpansive mappings in Banach spaces,” J. Math. Anal. Appl., vol. 67, no. 2, pp. 274–276, 1979.
[10] N. Nedic and A. Ozdaglar, “Distributed subgradient methods for multiagent optimization,” IEEE Trans. Autom. Control, vol. 54, no. 1, pp. 48–61, Jan. 2009.
[11] S. Liu, Z. Qiu, and L. Xie, “Convergence rate analysis of distributed optimization with projected subgradient algorithm,” Automatica, vol. 83, pp. 162–169, 2017.
[12] X. Li, L. Xie, and Y. Hong, “Distributed continuous-time nonsmooth convex optimization with coupled inequality constraints,” IEEE Trans. Control Netw. Syst., vol. 7, no. 1, pp. 74–84, Mar. 2020.

[13] V. G. L. Mejia, F. L. Lewis, Y. Wan, E. N. Sanchez, and L. Fan, “Solutions for multi-agent pursuit-evasion games on communication graphs: Finite-time capture and asymptotic behaviors,” IEEE Trans. Autom. Control, vol. 65, no. 5, pp. 1911–1923, May 2020.

[14] W. Ren and R. W. Beard, Distributed Consensus in Multi-Vehicle Cooperative Control. London, U.K.: Springer-Verlag, 2008.

[15] X. Li, M. Z. Q. Chen, and H. Su, “Quantized consensus of multi-agent networks with quantized inputs and Markovian communication links,” IEEE Trans. Cybern., vol. 49, no. 5, pp. 1816–1825, May 2019.

[16] D. Fullmer and A. S. Morse, “A distributed algorithm for computing a common fixed point of a finite family of paracontractions,” IEEE Trans. Autom. Control, vol. 63, no. 9, pp. 2833–2843, Sep. 2018.

[17] D. Fullmer, J. Liu, and A. S. Morse, “An asynchronous distributed algorithm for computing a common fixed point of a family of paracocontractions,” in Proc. 55th Conf. Decis. Control, Las Vegas, USA, 2016, pp. 2620–2625.

[18] J. Liu, D. Fullmer, A. Nedić, T. Başar, and A. S. Morse, “A distributed algorithm for computing a common fixed point of a family of strongly quasi-nonexpansive maps,” in Proc. Amer. Control Conf., Seattle, USA, 2017, pp. 686–690.

[19] S. S. Alaviani and N. Elia, “Distributed multi-agent convex optimization over random digraphs,” IEEE Trans. Autom. Control, vol. 65, no. 3, pp. 986–998, Mar. 2020.

[20] X. Li and G. Feng, “Distributed algorithms for computing a common fixed point of a group of nonexpansive operators,” IEEE Trans. Autom. Control, vol. 66, no. 5, pp. 2130–2145, May 2021.

[21] X. Li and L. Xie, “Distributed algorithms for computing a fixed point of multi-agent nonexpansive operators,” Automatica, vol. 122, 2020, Art. no. 109286.

[22] I. Necoara, P. Richtárik, and A. Patrascu, “Randomized projection methods for convex feasibility problems: Conditioning and convergence rates,” SIAM J. Optim., vol. 29, no. 4, pp. 2814–2852, 2019.

[23] A. Y. Kruger, D. R. Luke, and N. H. Thao, “Set regularities and feasibility of algorithms involving averaged nonexpansive operators,” in Proc. 55th Conf. Decis. Control, Las Vegas, USA, 2016, pp. 1–20.

[24] D. Fullmer and A. S. Morse, “A distributed algorithm for computing a common fixed point of a family of paracontractions,” in Proc. 55th Conf. Decis. Control, Las Vegas, USA, 2016, pp. 2620–2625.

[25] S. Mou, J. Liu, and A. S. Morse, “A distributed algorithm for solving a system of linear feasibility problems,” IEEE Trans. Control Netw. Syst., vol. 6, no. 1, pp. 82–93, Mar. 2019.

[26] S. S. Alaviani and N. Elia, “A distributed algorithm for solving linear algebraic equations over random networks,” in Proc. 57th Conf. Decis. Control, Miami Beach, FL, USA, 2018, pp. 83–88.

[27] X. Li, M. Meng, and L. Xie, “A linearly convergent algorithm for multi-agent quasi-nonexpansive operators in reflexive Banach spaces,” in Proc. 59th IEEE Conf. Decis. Control, Jeju Island, Korea, 2020, pp. 4903–4908.

[28] H. H. Bauschke and J. M. Borwein, “On projection algorithms for solving convex feasibility problems,” SIAM Rev., vol. 38, no. 3, pp. 367–426, 1996.

[29] H. H. Bauschke, D. Noll, and H. M. Phan, “Linear and strong convergence of algorithms involving averaged nonexpansive operators,” J. Math. Anal. Appl., vol. 421, no. 1, pp. 1–20, 2015.

[30] G. Banjac and P. J. Goulart, “Tight global linear convergence rate bounds for operator splitting methods,” IEEE Trans. Autom. Control, vol. 63, no. 12, pp. 4126–4139, Dec. 2018.

[31] A. Cegielski, S. Reich, and R. Zalas, “Regular sequences of quasi-nonexpansive maps,” in Proc. 57th Conf. Decis. Control, Miami Beach, FL, USA, 2018, pp. 1508–1532, 2018.

[32] L. Debath and P. Mikusinski, Introduction to Hilbert Spaces With Applications. Amsterdam, The Netherlands: Academic Press, 2005.

[33] M. Bianchi and S. Grammatico, “Fully distributed Nash equilibrium seeking over time-varying communication networks with linear convergence rate,” IEEE Control Syst. Lett., vol. 5, no. 2, pp. 499–504, Apr. 2021.

[34] R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction, 2nd ed. MIT Press, 2018.

[35] R. Xin, A. K. Sahu, U. A. Khan, and S. Kar, “Distributed stochastic optimization with gradient tracking over strongly-connected networks,” in Proc. 58th Conf. Decis. Control, Nice, France, 2019, pp. 8335–8338.

[36] S. Matsushita, “On the convergence rate of the Krasnosel’ski–Mann iteration,” Bull. Australian Math. Soc., vol. 96, no. 1, pp. 162–170, 2017.

[37] J. Zhang and K. You, “Decentralized stochastic gradient tracking for non-convex empirical risk minimization,” 2019, arXiv:1909.02712.

[38] X. Li, X. Yi, and L. Xie, “Distributed online optimization for multi-agent networks with coupled inequality constraints,” IEEE Trans. Autom. Control, vol. 66, no. 8, pp. 4344–4348, Aug. 2021.

[39] T. Charalambous, M. G. Rabbat, M. Johansson, and C. N. Hadjicostis, “Distributed finite-time computation of digraph parameters: Left-eigenvector, out-degree and spectrum,” IEEE Trans. Control Netw. Syst., vol. 3, no. 2, pp. 137–148, Jun. 2016.

[40] A. Nedić, A. Olshevsky, and W. Shi, “Achieving geometric convergence for distributed optimization over varying graphs,” SIAM J. Optim., vol. 27, no. 4, pp. 2597–2633, 2017.

[41] J. Xu, S. Zhu, Y. C. Soh, and L. Xie, “Convergence of asynchronous distributed gradient methods over stochastic networks,” IEEE Trans. Autom. Control, vol. 63, no. 2, pp. 434–448, Feb. 2018.

[42] D. Jaković, J. M. F. Moura, and J. Xavier, “Linear convergence rate of a class of distributed augmented Lagrangian algorithms,” IEEE Trans. Autom. Control, vol. 60, no. 4, pp. 922–936, Apr. 2015.

[43] R. Xin and U. A. Khan, “Distributed heavy-ball: A generalization and acceleration of first-order methods with gradient tracking,” IEEE Trans. Autom. Control, vol. 65, no. 6, pp. 2627–2633, Jun. 2020.

[44] S. Pu, W. Shi, J. Xu, and A. Nedić, “Push-pull gradient methods for distributed optimization in networks,” IEEE Trans. Autom. Control, vol. 66, no. 1, pp. 1–16, Jan. 2021.

[45] X. Li, L. Xie, and Y. Hong, “Distributed aggregative optimization over multi-agent networks,” IEEE Trans. Autom. Control, vol. 67, no. 6, pp. 3165–3171, Jun. 2022.

[46] S. Liang, L. Y. Wang, and G. Yin, “Exponential convergence of distributed primal-dual convex optimization algorithm without strong convexity,” Automatica, vol. 105, pp. 298–306, 2019.

[47] 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, 2020.

[48] M. Bianchi, G. Belgioioso, and S. Grammatico, “A distributed proximal-point algorithm for Nash equilibrium seeking under partial-decision information with geometric convergence,” 2019, arXiv:1910.11613.

[49] M. Meng and X. Li, “On the linear convergence of distributed Nash equilibrium seeking for multi-cluster games under partial-decision information,” Automatica, vol. 151, 2023, Art. no. 110919.

[50] P.-W. Wang and C.-J. Lin, “Iteration complexity of feasible descent methods for convex optimization,” J. Mach. Learn. Res., vol. 15, no. 1, pp. 1523–1548, 2014.

[51] R. A. Horn and C. R. Johnson, Matrix Analysis, 2nd ed. New York, NY, USA: Cambridge Univ. Press, 2012.
Min Meng received the B.S. and Ph.D. degrees in applied mathematics and control theory from Shandong University, Jinan, China, in 2010 and 2015, respectively. From April to October in 2014, from July to September in 2016, and from January to March in 2017, she was a Research Associate with the Department of Mechanical Engineering, The University of Hong Kong, Hong Kong. From 2015 to 2016, she was a Research Associate with the Department of Biomedical Engineering, City University of Hong Kong, Hong Kong. From 2017 to 2020, she was a Research Fellow with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. In 2020, she joined Tongji University, Shanghai, China, where she is currently a Professor. Her research interests include multiagent systems, distributed games and optimization, Boolean networks, and distributed secure control and estimation.

Lihua Xie (Fellow, IEEE) received the Ph.D. degree in electrical engineering from the University of Newcastle, Newcastle, Australia, in 1992. Since 1992, he has been with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, where he is currently a Professor and Director of the Center for Advanced Robotics Technology Innovation. From 2011 to 2014, he was the Head of Division of Control and Instrumentation, and from 2016 to 2021, Co-Director of Delta-NTU Corporate Lab for Cyber-Physical Systems. From 1986 to 1989, he held teaching positions with the Department of Automatic Control, Nanjing University of Science and Technology, Nanjing, China. His research interests include robust control and estimation, networked control systems, multiagent networks, localization, and unmanned systems. Dr. Xie is an Editor-in-Chief for *Unmanned Systems* and has served as Editor of IET Book Series in Control and Associate Editor for a number of journals including *Automatica*, IEEE TRANSACTIONS ON AUTOMATIC CONTROL, IEEE TRANSACTIONS ON CONTROL SYSTEMS TECHNOLOGY, IEEE TRANSACTIONS ON CONTROL OF NETWORK SYSTEMS, and IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS-II. From 2012 to 2014, he was an IEEE Distinguished Lecturer. He is a Fellow of the Academy of Engineering Singapore, IFAC, and CAA.