Structural Solutions for Additively Coupled Sum Constrained Games

Yi Su and Mihaela van der Schaar

***Abstract*—We propose and analyze a new family of games played by resource-constrained players. In particular, each strategic user has a single sum resource constraint over its action space and its own action impacts its own payoff through additive combinations of the other users' actions. We investigate convergence properties of various solutions for these games with and without real-time information exchange. First, when users cannot exchange messages with each other, but desire to maximize their individual utilities, we derive sufficient conditions under which best response dynamics converges to a globally asymptotically stable Nash equilibrium. Second, when users can exchange price signals in real-time to achieve coordination, we also establish the convergence properties of two action update mechanisms, including gradient play and Jacobi update. The investigated game model and our proposed solutions are readily applicable to various multi-user interaction, including communication networking applications, such as power control and flow control.**

***Index Terms*—Game theory, multi-user communications, Nash equilibrium, best-response dynamics, Gradient play, Jacobi update, pricing mechanism.**

I. Introduction

GAME theory provides a formal framework for describing and analyzing the interactions of users that behave strategically. Recently, there has been a surge in research activities that adopt game theoretic tools to investigate a wide range of modern communications and networking problems, such as flow and congestion control, network routing, load balancing, power control, peer-to-peer content sharing, etc [1]- [5]. In resource-constrained communication networks, any action taken by a user usually affects the utilities of the other users sharing the same resources and hence, it needs to be carefully chosen. Depending on the characteristics of different applications, numerous game-theoretical models and solution concepts have been proposed to characterize the multi-user interactions and optimize the users' decisions in communication networks. A variety of game theoretic solutions have been developed to characterize the resulting performance of the multi-user interaction, including Nash equilibrium (NE) and Pareto optimality [6].

The majority of the existing game theoretic research works in communication networking applications usually depend on the specific structures of action sets and utility functions in the investigated multi-user interaction. By considering or even architecting these specific structures, the associated games become analytically tractable and possess various important convergence properties. For instance, if users cannot exchange messages with each other and choose to individually maximize their utilities, to show the existence of and the convergence to a pure NE, several well-investigated classes of game models and frameworks, such as concave games, supermodular games, potential games, and variational inequality theory, have been extensively applied in various communication scenarios [7]- [13]. When real-time information exchange is possible, various mechanisms have also been proposed to enable collaborative users to jointly improve their performance and find the optimum joint policy. A well-known example is the framework of network utility maximization (NUM) started by Kelly etc [14]. It has recently been widely adopted in communication networks to analyze the trade-off between fairness and efficiency and various distributed resource allocation algorithms has been designed within the framework. In particular, for a convex NUM problem that can be decomposed into several subproblems by introducing Lagrange multipliers associated with different resource constraints, its global optimum can be computed using distributed algorithms by deploying message passing mechanisms [15].

To our knowledge, power control is one of the first few communication problems in which researchers start to apply game theoretic tools to investigate various properties. An interesting and important topic that has been extensively addressed recently is how to optimize multiple devices' power allocation when sharing a common frequency-selective interference channel. In [18], Yu et. al. first defined such a power control game from a game-theoretic perspective, proposed a best-response algorithm in which all users iteratively update their power allocations using the water-filling solution, and proved several sufficient conditions under which the algorithm globally converge to a unique pure NE. Many follow-up papers further establish several weaker sufficient convergence conditions with or without real-time information exchange [19]- [23]. We note that, the class of power control games is, on the one hand, narrow, but on the other hand, very important for communications. This is because in resource-constrained communication systems, there exist a family of multi-user interaction scenarios that share similar structures as the power control games. The purpose of this paper is to introduce and analyze a general framework that abstracts the common characteristics of this family of multi-user interaction scenarios. We explore the specific structures of the induced coupling among users and derive sufficient conditions that guarantee the convergence of different generic distributed algorithms

Paper approved by A. MacKenzie, the Editor for Game Theory and Cognitive Networking of the IEEE Communications Society. Manuscript received May 25, 2011; revised March 19 and July 3, 2012.

The authors are with the Department of Electrical Engineering, UCLA (e-mail: {yisu, mihaela}@ee.ucla.edu).

This work was supported by NSF 0830556 and ONR.

Digital Object Identifier 10.1109/TCOMM.2012.091312.110334

given the availability of real-time information exchange. In particular, the main contributions of this paper are as follows.

First of all, we define the model of *Additively Coupled Sum Constrained Games* (ACSCG) that arises in several communication and networking application. In particular, ACSCG is a special case of Rosen's concave games [30] and the central features of ACSCG are: 1) each user is resource-constrained in the sense that each of them is individually subject to a sum constraint; 2) users' utilities are separable across different dimensions in their action spaces; 3) the payoff in each dimension is determined by an additive combination of its own action and a function of the other users' actions.

Second, we derive the convergence conditions of various generic distributed algorithms with and without real-time information exchange. When no message exchanges between users is available and every user behaves to maximize its own utility, whether a NE exist and how to achieve it are of particular interests. In ACSCG, a pure NE exists in ACSCG due to Rosen's result. Our key contribution in this context is that we consider the best response dynamics to search for such a pure NE. We explore the properties of the additive coupling among users given the sum constraint and provide several sufficient conditions under which best response dynamics converges linearly[1] to the unique NE of ACSCG, for any set of feasible initialization with either sequential or parallel updates. We also explain the relationship between our results and the conditions previously developed in the game theory literature [30]- [32]. For games in which every user's action space is a vector set subjected to a single sum-constraint, our results identify sufficient conditions that guarantee both best response and gradient play dynamics globally converge to a pure NE, which can be regarded as a counterpart of Gabay's dominance solvability condition in games with single dimensional strategy [32]. When users can collaboratively exchange messages to coordinate with each other, we present the sufficient convergence conditions of two alternative distributed pricing algorithms, i.e. gradient play and Jacobi update. Our proposed convergence conditions generalize the results that have been previously obtained in [18]- [23] for the multi-user power control problem and they are immediately applicable to other multi-agent applications in communication networks that fulfill the requirements of ACSCG.

The rest of this paper is organized as follows. Section II defines the model of ACSCG. For ACSCG models, Sections III and IV present several distributed algorithms with and without real-time information exchange respectively and provide sufficient conditions that guarantee the convergence of the proposed algorithms. Section V presents the numerical examples and conclusions are drawn in Section VI.

II. Game Model

In this section, we introduce some basic definitions from the theory of strategic games, define the model of ACSCG, and present some illustrative examples.

[1]A sequence $x^{(k)}$ with limit x^* is linearly convergent if there exists a constant $c \in (0, 1)$ such that $|x^{(k)} - x^*| \le c|x^{(k-1)} - x^*|$ for k sufficiently large [29].

A. Strategic Games, Nash equilibrium, and Pareto Optimality

A strategic game is a suitable model for the analysis of a game where all users act independently and simultaneously according to their own self-interests and with no or limited a priori knowledge of the other users' strategies. This can be formally defined as a tuple $\Gamma = \langle \mathcal{N}, \mathcal{A}, u \rangle$. In particular, $\mathcal{N} = \{1, 2, \ldots, N\}$ is the set of rational decision-makers. Define $\mathcal{A}$ to be the joint action set $\mathcal{A} = \times_{n\in\mathcal{N}} \mathcal{A}_n$, with $\mathcal{A}_n$ being the action set available for user n. The vector utility function $u = \times_{n\in\mathcal{N}} u_n$ is a mapping from the individual users' joint action set to real numbers, i.e. $u : \mathcal{A} \rightarrow \mathcal{R}^N$. In particular, $u_n(\mathbf{a}) : \mathcal{A} \rightarrow \mathcal{R}$ is the utility of the nth user that generally depends on the strategies $\mathbf{a} = (\mathbf{a}_n, \mathbf{a}_{-n})$ of all users, where $\mathbf{a}_n \in \mathcal{A}_n$ denotes a feasible action profile of user n, and $\mathbf{a}_{-n} = \times_{m\neq n} \mathbf{a}_m$ is a vector of the strategies of all users except n. We also denote by $\mathcal{A}_{-n} = \times_{m\neq n} \mathcal{A}_m$ the joint action set of all users except n. To capture the multi-user performance tradeoff, the utility region is defined as $\mathcal{U} = \{(u_1(\mathbf{a}), \ldots, u_N(\mathbf{a}))| \ \exists \ \mathbf{a} \in \mathcal{A}\}$. Various game theoretic solutions were developed to characterize the resulting performance in both models, among which the most well-known ones include NE and Pareto optimality [6]. Significant research efforts have been devoted in the literature to constructing operational algorithms in order to achieve NE and Pareto optimality in various games with special structures of action set $\mathcal{A}_n$ and utility function u_n.

1) Nash equilibrium: definition, existence, and convergence: To avoid the overhead of performing real-time information exchange, network designers may prefer fully decentralized solutions in which the participating users simply compete against other users by choosing actions $\mathbf{a}_n \in \mathcal{A}_n$ to selfishly maximize their individual utility functions $u_n(\mathbf{a}_n, \mathbf{a}_{-n})$, given the actions $\mathbf{a}_{-n} \in \mathcal{A}_{-n}$. Most of these approaches focus on investigating the existence and properties of NE. NE is defined to be an action profile $(\mathbf{a}_1^*, \mathbf{a}_2^*, \ldots, \mathbf{a}_N^*)$ with the property that for every player, it satisfies $u_n(\mathbf{a}_n^*, \mathbf{a}_{-n}^*) \ge u_n(\mathbf{a}_n, \mathbf{a}_{-n}^*)$ for all $\mathbf{a}_n \in \mathcal{A}_n$, i.e. given the other users' actions, no user can increase its utility alone by changing its action. For an extensive discussion of the methodologies studying the existence, uniqueness, and convergence of various equilibrium in communication networks, we refer the readers to [16]. Many of the well-known results rely on specific structural properties of action set $\mathcal{A}$ and utility function u in the investigated multi-user interactions. For example, to establish the existence of and convergence to a pure NE, we can examine whether $\mathcal{A}$ and u satisfy the conditions of concave games, supermodular game, potential game, etc. Specifically, to apply the existence result of a pure NE in concave games [30], we need to check the following conditions: i) each player's action set $\mathcal{A}_n$ is convex and compact; and ii) the utility function $u_n(\mathbf{a}_n, \mathbf{a}_{-n})$ is continuous in $\mathbf{a}$ and quasi-concave[2] in $\mathbf{a}_n$ for any fixed $\mathbf{a}_{-n}$. As additional examples of games that guarantee the convergence to NE, it is well-known that, in supermodular games [7] [9] and potential games [11] [12], the best response dynamics can be used to search for a pure NE. Suppose that utility function u_n is twice continuously

[2]$f : \mathrm{R}^n \rightarrow \mathrm{R}$ is quasi-concave if $\mathrm{dom} f$ is convex and $\{x \in \mathrm{dom} f | f(x) \ge \alpha\}$ are convex for all α.

differentiable, $\forall n \in \mathcal{N}$. If $\mathcal{A}_n$ is a compact subset of $\mathcal{R}$ (or more generally $\mathcal{A}_n$ is a nonempty and compact sublattice[3]), $\forall n \in \mathcal{N}$, establishing that game Γ is a supermodular game is equivalent to showing that u_n satisfies

$$\forall (m,n) \in \mathcal{N}^2, m \neq n, \frac{\partial^2 u_n}{\partial \mathbf{a}_n \partial \mathbf{a}_m} \geq 0. \quad (1)$$

If action set $\mathcal{A}$ in game Γ is an interval of real numbers, we can show that game Γ is a potential game by verifying

$$\forall (m,n) \in \mathcal{N}^2, m \neq n, \frac{\partial^2 (u_n - u_m)}{\partial \mathbf{a}_n \partial \mathbf{a}_m} = 0. \quad (2)$$

2) Pareto optimality and network utility maximization: It is important to note that operating according to a Nash strategy will generally limit the performance of the user itself as well as of the entire network, because the available network resources are not always effectively exploited due to the conflicts of interest occurring among users. As opposed to the NE-based approaches, there exists a large body of literature that focuses on studying how users can *jointly* improve the system performance by optimizing a certain common objective function $f(u_1(\mathbf{a}), u_2(\mathbf{a}), \ldots, u_N(\mathbf{a}))$. This function represents the allocation rule based on which the system-wide resource allocation is performed. Different objective functions, e.g. sum utility maximization in which $f(u_1(\mathbf{a}), u_2(\mathbf{a}), \ldots, u_N(\mathbf{a})) = \sum_{n=1}^{N} u_n(\mathbf{a})$, can provide reasonable allocation outcomes by considering the trade-off between fairness and efficiency. A profile of actions is Pareto optimal if there is no other profile of actions that makes every user at least as well off and at least one user strictly better off.

The majority of these approaches focus on studying how to efficiently or distributively find the optimum joint policy. There exists a large body of literature that investigates how to compute Pareto optimal solutions in large-scale networks where centralized solutions are infeasible. Many structural results have been obtained for many generic distributed algorithms. A famous example is the NUM framework that develops distributed algorithms to solve network resource allocation problems [14]. The majority of the results in the existing NUM literature are based on convex optimization theory, in which the investigated problems share the following structures: the objective function $f(u_1(\mathbf{a}), u_2(\mathbf{a}), \ldots, u_N(\mathbf{a}))$ is convex[4], inequality resource constraint functions are convex, and equality resource constraint functions are affine. It is well-known that, for convex optimization problems, users can collaboratively exchange price signals that reflect the "cost" for consuming the constrained resources and the Pareto optimal allocations that maximizes the network utility can be determined in a fully distributed manner [15].

Summarizing, these general structural results with and without real-time message exchange turn out to be very useful when analyzing various multi-user interactions in communication networks. A lot of existing works are devoted to constructing or shaping the multi-user coupling such that it fits into these frameworks and the corresponding generic solutions can be directly applied. In the remaining part of this paper, we will derive several structural results for a particular type of multi-user interaction scenario. The following subsection gives the definition of ACSCG and presents several exemplary application scenarios.

B. Additively Coupled Sum Constrained Games

Definition 1: A multi-user interaction $\Gamma = \langle \mathcal{N}, \mathcal{A}, u \rangle$ is considered a ACSCG if it satisfies the following assumptions:

A1: $\forall n \in \mathcal{N}$, action set $\mathcal{A}_n \subseteq \mathcal{R}^K$ is defined to be[5]

$$\mathcal{A}_n = \Big\{(a_n^1, a_n^2, \cdots, a_n^K) \mid a_n^k \in [a_{n,k}^{\min}, a_{n,k}^{\max}] \text{ and } \sum_{k=1}^{K} a_n^k \leq M_n\Big\}. \quad (3)$$

A2: The utility function u_n satisfies

$$u_n(\mathbf{a}) = \sum_{k=1}^{K} \Big[h_n^k\big(a_n^k + f_n^k(\mathbf{a}_{-n})\big) - g_n^k(\mathbf{a}_{-n})\Big], \quad (4)$$

in which $h_n^k(\cdot)$ is an increasing and strictly concave function. Functions $f_n^k(\cdot)$ and $g_n^k(\cdot)$ are both twice differentiable.

The ACSCG model defined by assumptions A1 and A2 covers a broad class of multi-user interactions. Assumption A1 indicates that each player's action set is a K-dimensional vector set and its action vector is sum-constrained. This represents the communication scenarios in which each user needs to determine its multidimensional action in various channels or networks while the total amount of resources it can consume is constrained. Assumption A2 implies that a user's utility is separable and can be represented by the summation of concave functions h_n^k minus "penalty" functions g_n^k across the K dimensions. In particular, within each dimension, the input of h_n^k is an additive combination of a_n^k and $f_n^k(\mathbf{a}_{-n})$. Since a_n^k only appears in the concave function h_n^k, it implies that each user's utility is concave in its own action, i.e. diminishing returns per unit of user n's invested action $\mathbf{a}_n$, which is common for many application scenarios in communication networks. The key features of the game model defined by A1 and A2 include: each user's action is subjected to a *sum constraint*; users' utilities are impacted by *additive combinations* of a_n^k and $f_n^k(\mathbf{a}_{-n})$ through concave functions h_n^k. Therefore, we term game Γ that satisfies assumptions A1 and A2 as ACSCG. In the following section, we present several illustrative multi-user interaction examples that belong to ACSCG.

C. Examples of ACSCG

Example 1: We first consider a simple two-user game with two-dimension action spaces, i.e. $N = K = 2$. The utility functions are given by[6]

$$u_n(\mathbf{a}) = \sqrt{a_n^1 + \frac{(a_{-n}^1)^2}{4} - \frac{(a_{-n}^2)^2}{9}} + \sqrt{a_n^2 - \frac{(a_{-n}^1)^2}{5} + \frac{(a_{-n}^2)^2}{2}}$$

[3]A real m-dimensional set $\mathcal{V}$ is a *sublattice* of $\mathcal{R}^m$ if for any two elements $a, b \in \mathcal{V}$, the component-wise minimum, $a \wedge b$, and the component-wise maximum, $a \vee b$, are also in $\mathcal{V}$.

[4]$f: \mathrm{R}^n \rightarrow \mathrm{R}$ is convex if dom$f$ is a convex set and $f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y)$, $\forall x, y \in \mathrm{dom} f, 0 \leq \theta \leq 1$.

[5]We consider a sum constraint throughout the paper rather than a weighted-sum constraint, because a weighted-sum constraint can be easily converted to a sum constraint by rescaling $\mathcal{A}_n$. Besides, we nontrivially assume that $\sum_{k=1}^{K} a_{n,k}^{\max} \geq M_n$.

[6]In this example, since there are only two users, the subindex $-n$ denotes the user but n.

for $n = 1, 2$. The resource constraints are $\sum_{k=1}^{2} a_n^k \leq M_n$ in which $M_n > 0$ and $a_n^k \geq 0$ for $\forall n, k$.

Example 2: (Power control in frequency-selective Gaussian interference channel [18] [21]) There are N transmitter and receiver pairs in the system. The entire frequency band is divided into K frequency bins. In frequency bin k, the channel gain from transmitter i to receiver j is denoted as H_{ij}^k, where $k = 1, 2, \cdots, K$. Similarly, denote the noise power spectral density (PSD) that receiver n experiences as σ_n^k and player n's transmit PSD as P_n^k. The action of user n is to select its transmit power $\mathbf{P}_n = [P_n^1 \; P_n^2 \cdots P_n^K]$ and the transmit PSD is subject to its power constraint: $\sum_{k=1}^{K} P_n^k \leq \mathrm{P}_n^{\max}$. For a fixed $\mathbf{P}_n$, if treating interference as noise, user n can achieve the data rate in (5).

Example 3: (*Delay minimization in Jackson Networks [17]*) As an additional example, we consider a network of N nodes. A Poisson stream of external packets arrive at node n with rate ψ_n and the input stream is split into K traffic classes, which are individually served by exponential servers. Denote node n's input rate and service rate for class k as ψ_n^k and μ_n^k respectively. Therefore, the action of node n is to determine the rates for different traffic classes $\Psi_n = [\psi_n^1 \; \psi_n^2 \cdots \psi_n^K]$ and the total rate is subject to the minimum rate constraint: $\sum_{k=1}^{K} \psi_n^k \geq \psi_n^{\min}$. The packets of the same traffic class constitute a Jackson network in which Markovian routing is adopted: packets of class k completing service at node m are routed to node n with probability r_{mn}^k or exit the network with probability $r_{m0}^k = 1 - \sum_{n=1}^{N} r_{mn}^k$. Denote the arrival rate for class k at node n as η_n^k. By Jackson's Theorem, we have $\eta_n^k = \psi_n^k + \sum_{m=1}^{N} \eta_m^k r_{mn}^k, n = 1, 2, \cdots, K$. Denote $[\mathrm{R}^k]_{mn} = r_{nm}^k$, $\Upsilon^k = (\mathrm{I} - \mathrm{R}^k)^{-1}$, and $\upsilon_{mn}^k = [\Upsilon^k]_{nm}$. Equivalently, we have $\eta_n^k = \sum_{m=1}^{N} \upsilon_{mn}^k \psi_m^k$. Each node aims to minimize its total additional M/M/1 queueing delay incurred by accommodating its traffic:

$$d_n(\mathbf{\Psi}) = \sum_{k=1}^{K} \Big(\frac{1}{\mu_n^k - \sum_{m=1}^{N} \upsilon_{mn}^k \psi_m^k} - \frac{1}{\mu_n^k - \sum_{m \neq n} \upsilon_{mn}^k \psi_m^k} \Big). \tag{6}$$

Example 4: (Asynchronous transmission in digital subscriber lines network [20]) The basic setting of this example is similar as that of Example 2 except that inter-carrier interference (ICI) exist among different frequency bins. Due to the loss of the orthogonality, the interference that user n experiences in frequency bin k is

$$f_n^k(\mathbf{P}_{-n}) = \sum_{m \neq n} \Big(\sum_{j=1}^{K} \gamma(k-j) H_{mn}^j P_m^j \Big), \tag{7}$$

in which $\gamma(j)$ is the ICI coefficients that represents the relative interference transmitted signal in a particular frequency bin generates to its jth neighbor bin. In particular, it takes the form

$$\gamma(j) = \begin{cases} 1, & \text{if } j = 0 \\ \frac{2}{K^2 \sin^2(\frac{\pi}{K} j)}, & -\frac{K}{2} \leq j \leq \frac{K}{2} \; j \neq 0. \end{cases} \tag{8}$$

It satisfies the symmetric and circular properties, i.e. $\gamma(-j) = \gamma(j) = \gamma(K-j)$. User n's achievable rate in the presence of ICI is given by

$$r_n(\mathbf{P}) = \sum_{k=1}^{K} \log_2 \left[1 + \frac{H_{nn}^k P_n^k}{\sigma_n^k + \sum_{m \neq n} \Big(\sum_{j=1}^{K} \gamma(k-j) H_{mn}^j P_m^j \Big)} \right]. \tag{9}$$

We can verify that Examples 1-4 satisfy assumptions A1 and A2 and belong to ACSCG. The details of functions $h_n^k(\cdot)$, $f_n^k(\cdot)$ and $g_n^k(\cdot)$ are summarized in Table I. For each example, Table I also summarizes the applicable convergence conditions that will be provided in the remaining parts of the paper. We would like to mention that Example 3 can be shown to be a special case of ACSCG by slightly transforming the action sets and utilities. We can define user n's action as $-\Psi_n$. For user n, the sum constraint becomes $\sum_{k=1}^{K} -\psi_n^k \leq -\psi_n^{\min}$ and minimizing $d_n(\mathbf{\Psi})$ is equivalent to maximizing $-d_n(\mathbf{\Psi})$.

Remark 1: (Issues related to ACSCG) Since the ACSCG model represents a good abstraction of numerous multi-user resource allocation problems, we aim to investigate the convergence properties of various distributed algorithms in ACSCG with and without real-time message passing.

On one hand, it is straightforward to verify that ACSCG is a special case of Rosen's concave games [30]. Therefore, at least one pure NE is admitted. In practice, we want to provide the sufficient conditions under which the best response dynamics provably and globally converges to a pure NE. However, the existing literature, e.g. the diagonal strict concavity (DSC) conditions in [30] and the supermodular game theory [7]- [9], does not provide such convergence conditions for the general ACSCG model. For example, the DSC conditions developed by Rosen for general concave games do not guarantee the convergence of best response dynamics [30]. Even if the utility functions in ACSCG possess the supermodular type structure, due to the sum constraint, the action set of each user is generally not a sublattice[7]. Therefore, the convergence results based on supermodular games cannot be directly applied in ACSCG. On the other hand, if we want to maximize the sum utility by enabling real-time message passing among users, we also note that, the utility u_n is not necessarily jointly convex in **a** because of the existence of $g_n^k(\cdot)$. Therefore, the existing algorithms developed for the convex NUM are not be immediately applicable either.

In fact, a unique feature of the ACSCG is that different users' actions are *additively coupled* in $h_n^k(\cdot)$ and each user's action space is *sum-constrained*. In the following sections, we will fully explore these specific structures and address the convergence properties of various distributed algorithms with and without real-time information exchange.

III. Solutions without Message Passing

In the applications where system designers focus on fully decentralized strategies to avoid the heavy signaling required to achieve coordination, the participating users can simply choose actions to selfishly maximize their individual utility functions $u_n(\mathbf{a})$ without taking into account the utility degredation caused to the other users. In particular, each user

[7] In supermodular games, for each player, the action set is a nonempty and compact sublattice. We can verify that with the sum constraint, $\mathcal{A}_n$ is usually not a sublattice by taking the component-wise maximum.

$$r_n(\mathbf{P}) = \sum_{k=1}^{K} \log_2 \Big(1 + \frac{H_{nn}^k P_n^k}{\sigma_n^k + \sum_{m \neq n} H_{mn}^k P_m^k}\Big) = \sum_{k=1}^{K} \Big(\log_2(\sigma_n^k + \sum_{m=1}^{N} H_{mn}^k P_m^k) - \log_2(\sigma_n^k + \sum_{m \neq n} H_{mn}^k P_m^k) \Big). \quad (5)$$

TABLE I
EXAMPLES 1-4 AS ACSCG.

Examples	$f_n^k(\mathbf{a}_{-n})$	$h_n^k(x)$	$g_n^k(\mathbf{a}_{-n})$	Convergence conditions
Example 1	$f_n^1(\mathbf{a}_{-n}) = \frac{(a_{-n}^1)^2}{4} - \frac{(a_{-n}^2)^2}{9}$ $f_n^2(\mathbf{a}_{-n}) = \frac{(a_{-n}^2)^2}{2} - \frac{(a_{-n}^1)^2}{5}$	$\sqrt{x}$	0	(C4) and (C6)
Example 2	$\sum_{m \neq n} \frac{H_{mn}^k}{H_{nn}^k} P_m^k$	$\log_2(\sigma_n^k + H_{nn}^k x)$	$\log_2(\sigma_n^k + \sum_{m \neq n} H_{mn}^k P_m^k)$	(C1)-(C8)
Example 3	$\sum_{m \neq n} \frac{\upsilon_{mn}^k}{\upsilon_{nn}^k} \psi_m^k$	$-\frac{1}{\mu_n^k - \upsilon_{nn}^k x}$	$-\frac{1}{\mu_n^k - \sum_{m \neq n} \upsilon_{mn}^k \psi_m^k}$	(C1)-(C8)
Example 4	$\sum_{m \neq n} (\sum_{j=1}^{K} \frac{\gamma(k-j) H_{mn}^k}{H_{nn}^k} P_m^k)$	$\log_2(\sigma_n^k + H_{nn}^k x)$	$\log_2\big(\sigma_n^k + H_{nn}^k f_n^k(\mathbf{a}_{-n})\big)$	(C4)-(C8)

individually solves the following optimization program:

$$\max_{\mathbf{a}_n \in \mathcal{A}_n} u_n(\mathbf{a}). \quad (10)$$

The steady state outcome of such a multi-user interaction is usually characterized as a NE, at which given the other users actions $\mathbf{a}_{-n}$, no user can increase its utility alone by unilaterally changing its action. It is worth pointing out that, since there is no coordination signal among users, the Nash strategy generally does not lead to a Pareto-optimal solution. Section IV will discuss distributed algorithms in which users exchange coordination signals in order to improve the system efficiency.

A. Properties of Best Response Dynamics in ACSCG

For ease of presentation, in this subsection, we temporarily focus on the scenarios in which $f_n^k(\mathbf{a}_{-n})$ takes the following form

$$f_n^k(\mathbf{a}_{-n}) = \sum_{m \neq n} F_{mn}^k a_m^k, \quad (11)$$

in which $F_{mn}^k \in \mathcal{R}$, $\forall m, n, k$. We can see that, for user n, $f_n^k(\mathbf{a}_{-n})$ is the linear combination of the remaining users' action in the same dimension k. Specifically, both Example 2 and 3 in Table I belong to this category. In Section III-B, we will extend the results derived for the functions $f_n^k(\mathbf{a}_{-n})$ defined in (11) to general $f_n^k(\mathbf{a}_{-n})$. The key differences between all the sufficient conditions that will be provided in this section are summarized in Table II.

Since $h_n^k(\cdot)$ is concave, the objective in (10) is a concave function when the other users' actions $\mathbf{a}_{-n}$ are fixed. To find the globally optimal solution of the problem in (10), we can first form its Lagrangian

$$L_n(\mathbf{a}_n, \lambda) = u_n(\mathbf{a}) + \lambda(M_n - \sum_{k=1}^{K} a_n^k), \quad (12)$$

in which $a_n^k \in [a_{n,k}^{\min}, a_{n,k}^{\max}]$. Take the first derivatives of (12), we have

$$\frac{\partial L_n(\mathbf{a}_n, \lambda)}{\partial a_n^k} = \frac{\partial h_n^k(a_n^k + \sum_{m \neq n} F_{mn}^k a_m^k)}{\partial a_n^k} - \lambda = 0. \quad (13)$$

Denote

$$l_n^k(\mathbf{a}_{-n}, \lambda) \triangleq \Big[\Big\{ \frac{\partial h_n^k}{\partial x} \Big\}^{-1}(\lambda) - \sum_{m \neq n} F_{mn}^k a_m^k \Big]_{a_{n,k}^{\min}}^{a_{n,k}^{\max}}, \quad (14)$$

in which $[x]_b^a = \max\{\min\{x, a\}, b\}$. The optimal solution of (10) is given by $a_n^{*k} = l_n^k(\mathbf{a}_{-n}, \lambda^*)$, where the Lagrange multiplier λ^* is chosen to satisfy the sum constraint $\sum_{k=1}^{K} a_n^{*k} = M_n$. We define the best response operator $B_n^k(\cdot)$ as

$$B_n^k(\mathbf{a}_{-n}) = l_n^k(\mathbf{a}_{-n}, \lambda^*). \quad (15)$$

We consider the dynamic adjustment process in which users revise their actions over time based on their observations about their opponents. A well-known candidate for such adjustment processes is the so-called best response dynamics. In the best response algorithm, each user updates its action using the best response strategy that maximizes its utility function in (4). We consider two types of update orders, including sequential update and parallel update. Specifically, in sequential update, individual players iteratively optimize in a circular fashion with respect to its own action while keeping the actions of its opponents fixed. Formally, at stage t, user n chooses its action according to

$$a_n^{k,t} = B_n^k([\mathbf{a}_1^t, \ldots, \mathbf{a}_{n-1}^t, \mathbf{a}_{n+1}^{t-1}, \ldots, \mathbf{a}_N^{t-1}]). \quad (16)$$

On the other hand, players adopting the parallel update their actions at stage t according to

$$a_n^{k,t} = B_n^k(\mathbf{a}_{-n}^{t-1}). \quad (17)$$

We obtain several sufficient conditions under which best response dynamics converges. Similar convergence conditions are proved in [19]- [21] for Example 2 in which $h_n^k(x) = \log_2(\sigma_n^k + H_{nn}^k x)$. We consider more general functions $h_n^k(\cdot)$ and further extend the convergence conditions in [19]- [21].

1) General $h_n^k(\cdot)$: The first sufficient condition is developed for the general cases in which the functions $h_n^k(\cdot)$ in the utilities $u_n(\cdot)$ are specified in assumption A2. Define

$$[\mathbf{T}^{\max}]_{mn} \triangleq \begin{cases} \max_k |F_{mn}^k|, & \text{if } m \neq n \\ 0, & \text{otherwise.} \end{cases} \quad (18)$$

and let $\rho(\mathbf{T}^{\max})$ denote the spectral radius of the matrix $\mathbf{T}^{\max}$.

Theorem 1: If

$$\rho(\mathbf{T}^{\max}) < \frac{1}{2}, \quad \text{(C1)}$$

TABLE II
COMPARISON AMONG CONDITIONS (C1)-(C6).

Conditions	Assumptions about $f_n^k(\mathbf{a}_{-n})$	$h_n^k(x)$	Measure of residual error $\mathbf{a}_n^{t+1} - \mathbf{a}_n^t$	Contraction factor
(C1)	(11)	A2	1-norm	$2\rho(\mathbf{T}^{\max})$
(C2)	(11) and F_{mn}^k have the same sign for $\forall k, m \neq n$	A2	1-norm	$\rho(\mathbf{T}^{\max})$
(C3)	(11)	(20)	weighted Euclidean norm	$\rho(\mathbf{S}^{\max})$
(C4)	general	A2	1-norm	$2\rho(\bar{\mathbf{T}}^{\max})$
(C5)	$\frac{\partial f_n^k(\mathbf{a}_{-n})}{\partial a_m^{k'}}$ have the same sign for $\forall \mathbf{a} \in \mathcal{A}, k, k', m \neq n$	A2	1-norm	$\rho(\bar{\mathbf{T}}^{\max})$
(C6)	general	(20)	weighted Euclidean norm	$\rho(\bar{\mathbf{S}}^{\max})$

then best response dynamics converges linearly to the unique NE of game Γ, for any set of initial conditions belonging to $\mathcal{A}$ with either sequential or parallel updates.

Proof: It can be proved by showing that the best response dynamics defined in (16) and (17) is a contraction mapping. See Appendix A for details. ■

In multi-user communication applications, it is common to have games of *strategic complements* (or *strategic substitutes*), i.e. the marginal returns to any one component of the player's action rise with increases (or decreases) in the components of the competitors' actions [35]. Mathematically, if u_n is twice differentiable, strategic complementarities (or strategic substitutes) can be described as

$$\frac{\partial^2 u_n(\mathbf{a}_n, \mathbf{a}_{-n})}{\partial a_n^j \partial a_m^k} \geq 0, \ \forall m \neq n, (\text{or } \frac{\partial^2 u_n(\mathbf{a}_n, \mathbf{a}_{-n})}{\partial a_n^j \partial a_m^k} \leq 0, \ \forall m \neq n). \tag{19}$$

For instance, in Examples 2 and 4, increasing user n's transmitted power creates stronger interference to the other users and decreases their marginal achievable rates. Similarly, in Example 3, increasing node n's input traffic rate congests all the servers in the network and increases the marginal queueing delay. For the ACSCG models that exhibit strategic complementarities (or strategic substitutes), the following theorem further relaxes condition (C1).

Theorem 2: For Γ with strategic complementarities (or strategic substitutes) in utility functions, i.e. $F_{mn}^k \leq 0$, $\forall k, m \neq n$, (or $F_{mn}^k \geq 0$, $\forall k, m \neq n$), if

$$\rho(\mathbf{T}^{\max}) < 1, \tag{C2}$$

then best response dynamics converges linearly to the unique NE of game Γ, for any set of initial conditions belonging to $\mathcal{A}$ with either sequential or parallel updates.

Proof: It can be shown by adapting the proof of Theorem 1. See Appendix B. ■

Remark 2: (Implications of conditions (C1) and (C2)) Theorem 1 and Theorem 2 give sufficient conditions for best response dynamics to globally converge to a unique fixed point. Specifically, $\max_k |F_{mn}^k|$ can be regarded as a measure of the strength of the mutual coupling between user m and n. The intuition behind (C1) and (C2) is that, the weaker the coupling among different users is, the more likely that best response dynamics converges. Consider the extreme case in which $F_{mn}^k = 0, \forall k, m \neq n$. Since each user's utility is not impacted by the remaining users' action $\mathbf{a}_{-n}$, the convergence is immediately achieved after a single best-response iteration. If no restriction is imposed on F_{mn}^k, Theorem 1 specifies a mutual coupling threshold under which best response dynamics provably converge. The proof of Theorem 1 can be intuitively interpreted as follows. We regard every best response update as the users' joint attempt to approach the NE. Due to the linear coupling structure in (11), user n's best response in (14) can be expressed in terms of linear combinations of $\mathbf{a}_{-n}$. As a result, the residual error $\left|\mathbf{a}_n^{t+1} - \mathbf{a}_n^t\right|_1$, which is the 1-norm distance between the updated action profile $\mathbf{a}_n^{t+1}$ and the current action profile $\mathbf{a}_n^t$, can be upper-bounded using linear combinations of $\left|\mathbf{a}_m^t - \mathbf{a}_m^{t-1}\right|_1$ in which $m \neq n$. Recall that F_{mn}^k can be either positive or negative. We also note that, if $\mathbf{a}_m^t \neq \mathbf{a}_m^{t-1}$, $\mathbf{a}_m^t - \mathbf{a}_m^{t-1}$ contains both positive and negative terms due to the sum-constraint. In the worst case, distance $\left|\mathbf{a}_n^{t+1} - \mathbf{a}_n^t\right|_1$ is maximized if $\left\{F_{mn}^k\right\}$ and $\left\{a_m^{k,t} - a_m^{k,t-1}\right\}$ are co-phase multiplied and additively summed, i.e. $F_{mn}^k\left(a_m^{k,t} - a_m^{k,t-1}\right) \geq 0$, for $\forall k = 1, \ldots, K, m \neq n$. After an iteration, all users except n contributes to user n's residual error at stage $t+1$ up to $\sum_{m \neq n} 2 \max_k \left|F_{mn}^k\right| \left|\mathbf{a}_m^t - \mathbf{a}_m^{t-1}\right|_1$. Under condition (C1), it is guaranteed that the residual error contracts with respect to the special norm defined in (70). Theorem 2 focuses on the situations in which the signs of F_{mn}^k are the same, $\forall m \neq n, k$. In this case, $\left\{F_{mn}^k\right\}$ and $\left\{a_m^{k,t} - a_m^{k,t-1}\right\}$ cannot be co-phase multiplied. Therefore, the range of convergence enlarges and hence, condition (C2) stated in Theorem 2 is weaker than condition (C1) in Theorem 1.

Remark 3: (Relation to the results in references [19]- [21]) Similarly as [19] [20], our proofs choose 1-norm as the distance measure for the residual errors $\mathbf{a}_n^{t+1} - \mathbf{a}_n^t$ after each best-response iteration. However, by manipulating the inequalities in a different way, condition (C2) is more general than the results in [19] [20], where they require $\max_k F_{mn}^k < \frac{1}{N-1}$. Interestingly, condition (C2) recovers the result obtained in [21] where it is proved by choosing Euclidean norm as the distance measure for the residual errors $\mathbf{a}_n^{t+1} - \mathbf{a}_n^t$ after each best-response iteration. However, the approach in [21] using Euclidean norm only applies to the scenarios in which $h_n^k(\cdot)$ is a logarithmic function. We prove that condition (C2) applies to any $h_n^k(\cdot)$ that is increasing and strictly concave.

2) A special class of $h_n^k(\cdot)$: In addition to conditions (C1) and (C2), we also develop a sufficient convergence condition for a family of utility functions parameterized by a negative number θ. In particular, $h_n^k(\cdot) : \mathcal{R}_+ \to \mathcal{R}_+$ satisfies

$$h_n^k(x) = \begin{cases} \log(\alpha_n^k + F_{nn}^k x), & \text{if } \theta = -1, \\ \frac{(\alpha_n^k + F_{nn}^k x)^{\theta+1}}{\theta+1}, & \text{if } -1 < \theta < 0 \text{ or } \theta < -1. \end{cases} \tag{20}$$

and $\alpha_n^k \in \mathcal{R}$ and $F_{nn}^k > 0$. The interpretation of this type of utilities has been addressed in [24]. It is shown that varying the parameter θ leads to different types of fairness across $\alpha_n^k + F_{nn}^k(a_n^k + \sum_{m \neq n} F_{mn}^k a_m^k)$ for all k. In particular, $\theta = -1$ corresponds to the proportional fairness; if $\theta = -2$, then harmonic mean fairness; and if $\theta = -\infty$, then max-min fairness. We can see that, Examples 2 and 3 are special cases of this type of utility functions. In these cases, best response dynamics in equation (14) is reduced to

$$l_n^k(\mathbf{a}_{-n}, \lambda) = \left[\left(\frac{1}{F_{nn}^k}\right)^{1+\frac{1}{\theta}} \lambda^{\frac{1}{\theta}} - \frac{\alpha_n^k}{F_{nn}^k} - \sum_{m \neq n} F_{mn}^k a_m^k \right]_{a_{n,k}^{\min}}^{a_{n,k}^{\max}}, \tag{21}$$

Define $[\mathbf{S}^{\max}]_{mn}$ in (22). For the class of utility functions in (20), Theorem 3 gives a sufficient condition that guarantees the convergence of the best response dynamics defined in (21).

Theorem 3: For $h_n^k(\cdot)$ defined in (20), if

$$\rho(\mathbf{S}^{\max}) < 1, \tag{C3}$$

then best response dynamics converges linearly to the unique NE of game Γ, for any set of initial conditions belonging to $\mathcal{A}$ and with either sequential or parallel updates.

Proof: It can be proved by showing that the best response dynamics defined in (21) is a contraction mapping with respect to weighted Euclidean norm. See Appendix C for details. ■

Remark 4: (Relation between conditions (C3) and the results in reference [21]) For aforementioned Example 2, Scutari et al. established in [21] a sufficient condition under which the iterative water-filling algorithm converges. The iterative water-filling algorithm is essentially belongs to best response dynamics. Specifically, in [21], Shannon's formula leads to $\theta = -1$ and cross channel coefficients satisfy $F_{mn}^k \geq 0, \forall k, m \neq n$. Equation (21) reduces to the water-filling formula

$$l_n^k(\mathbf{a}_{-n}, \lambda) = \left[\frac{1}{\lambda} - \frac{\alpha_n^k}{F_{nn}^k} - \sum_{m \neq n} F_{mn}^k a_m^k \right]_{a_{n,k}^{\min}}^{a_{n,k}^{\max}}, \tag{23}$$

and $[\mathbf{S}^{\max}]_{mn} = \max_k F_{mn}^k$. By choosing weighted Euclidean norm as the distance measure for the residual errors $\mathbf{a}_n^{t+1} - \mathbf{a}_n^t$ after each best-response iteration, Theorem 3 generalizes the results in [21] for the family of utility functions defined in (20).

Remark 5: (Relation between conditions (C1), (C2) and (C3)) The connections and differences between conditions (C1), (C2) and (C3) are summarized in Table II. We have addressed the implications of (C1) and (C2) in Remark 2. Now we discuss their relation with (C3). First of all, condition (C1) is proposed for general $h_n^k(\cdot)$ and condition (C3) is proposed for the class of utility functions defined in (20). However, Theorem 1 and Theorem 3 individually establish the fact that best response dynamics is a contraction map by selecting different vector and matrix norms. Therefore, in general, (C1) and (C3) does not immediately imply each other. Note that $[\mathbf{S}^{\max}]_{mn} \leq \zeta_{mn} \cdot \max_k |F_{mn}^k|$ in which ζ_{mn} satisfies (24). The physical interpretation of ζ_{mn} is the similarity between the preferences of user m and n across the total K dimensions of their action spaces. Recall that both $\mathbf{S}^{\max}$ and $\mathbf{T}^{\max}$ are non-negative matrices and $\mathbf{S}^{\max}$ is element-wise less than or equal to $\max_{m \neq n} \zeta_{mn} \mathbf{T}^{\max}$. By the property of non-negative matrix and condition (C1), we can conclude $\rho(\mathbf{S}^{\max}) \leq \rho(\max_{m \neq n} \zeta_{mn} \mathbf{T}^{\max}) < \max_{m \neq n} \frac{\zeta_{mn}}{2}$.

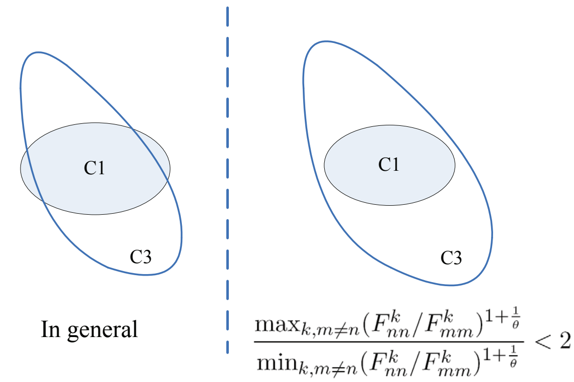

Fig. 1. Relation between (C1) and (C3).

The relation between (C1) and (C3) is pictorially illustrated in Fig. 1. Specifically, if users have similar preference in their available actions and the upper bound of ζ_{mn} that measures the difference of their preferences is below the following threshold:

$$\frac{\max_{k,m \neq n}(F_{nn}^k/F_{mm}^k)^{1+\frac{1}{\theta}}}{\min_{k,m \neq n}(F_{nn}^k/F_{mm}^k)^{1+\frac{1}{\theta}}} < 2, \tag{25}$$

we have $\rho(\mathbf{S}^{\max}) < 1$ and hence, (C1) is stronger than (C3). We also would like to point out that, the LHS of (25) is a function of θ and the LHS $\equiv 1$ if $\theta = -1$. When $\theta = -1$, $\mathbf{T}^{\max}$ coincides with $\mathbf{S}^{\max}$. Mathematically, in this case, (C3) is actually more general than (C2), because it still holds even if F_{mn}^k have different signs.

B. Extensions to General $f_n^k(\cdot)$

As a matter of fact, the results above can be extended to the more general situations in which $f_n^k(\cdot)$ is a nonlinear differentiable function, $\forall n, k$ and its input $\mathbf{a}_{-n}$ consists of the remaining users' action from all the dimensions. Accordingly, equation (14) becomes

$$l_n^k(\mathbf{a}_{-n}, \lambda) \triangleq \left[\left\{ \frac{\partial h_n^k}{\partial x} \right\}^{-1}(\lambda) - f_n^k(\mathbf{a}_{-n}) \right]_{a_{n,k}^{\min}}^{a_{n,k}^{\max}}. \tag{26}$$

The conclusions in Theorem 1, 2, and 3 can be further extended as Theorem 4, and 5, 6 that are listed below. We only provide the proof of Theorem 4 in Appendix D. The detailed proofs of Theorem 5 and 6 are omitted because they can be proven similarly as Theorem 4.

For general $f_n^k(\cdot)$, we denote

$$[\bar{\mathbf{T}}^{\max}]_{mn} \triangleq \begin{cases} \max_{\mathbf{a} \in \mathcal{A}, k'} \sum_{k=1}^K \left| \frac{\partial f_n^k(\mathbf{a}_{-n})}{\partial a_m^{k'}} \right|, & \text{if } m \neq n \\ 0, & \text{otherwise.} \end{cases} \tag{27}$$

Besides, for $h_n^k(\cdot)$ defined in (20), we define $[\bar{\mathbf{S}}^{\max}]_{mn}$ in (28).

Theorem 4: If

$$\rho(\bar{\mathbf{T}}^{\max}) < \frac{1}{2}, \tag{C4}$$

then best response dynamics converges linearly to the unique NE of game Γ, for any set of initial conditions belonging to $\mathcal{A}$ with either sequential or parallel updates.

$$[\mathbf{S}^{\max}]_{mn} \triangleq \begin{cases} \frac{\sum_{k=1}^K (F_{mm}^k)^{1+\frac{1}{\theta}}}{\sum_{k=1}^K (F_{nn}^k)^{1+\frac{1}{\theta}}} \max_k \left\{ |F_{mn}^k| \left(\frac{F_{nn}^k}{F_{mm}^k} \right)^{1+\frac{1}{\theta}} \right\}, & \text{if } m \neq n \\ 0, & \text{otherwise.} \end{cases} \quad (22)$$

$$\zeta_{mn} = \frac{\sum_{k=1}^K (F_{mm}^k)^{1+\frac{1}{\theta}}}{\sum_{k=1}^K (F_{nn}^k)^{1+\frac{1}{\theta}}} \cdot \max_k \frac{(F_{nn}^k)^{1+\frac{1}{\theta}}}{(F_{mm}^k)^{1+\frac{1}{\theta}}} \in \left[1, \frac{\max_k (F_{nn}^k / F_{mm}^k)^{1+\frac{1}{\theta}}}{\min_k (F_{nn}^k / F_{mm}^k)^{1+\frac{1}{\theta}}} \right]. \quad (24)$$

$$[\bar{\mathbf{S}}^{\max}]_{mn} \triangleq \begin{cases} \frac{\sum_{k=1}^K (F_{mm}^k)^{1+\frac{1}{\theta}}}{\sum_{k=1}^K (F_{nn}^k)^{1+\frac{1}{\theta}}} \max_{\mathbf{a} \in \mathcal{A}, k'} \left\{ \sum_{k=1}^K \left| \frac{\partial f_n^k(\mathbf{a}_{-n})}{\partial a_m^{k'}} \right| \left(\frac{F_{nn}^{k'}}{F_{mm}^{k'}} \right)^{1+\frac{1}{\theta}} \right\}, & \text{if } m \neq n \\ 0, & \text{otherwise.} \end{cases} \quad (28)$$

Proof: It can be proved by combining the proof of Theorem 1 and the mean value theorem for vector-valued functions. See Appendix D for details. ■

Similarly as in Theorem 2, for the general ACSCG models that exhibit strategic complementarities (or strategic substitutes), we can further relax condition (C4).

Theorem 5: For Γ with strategic complementarities (or strategic substitutes), i.e. $\frac{\partial f_n^k(\mathbf{a}_{-n})}{\partial a_m^{k'}} \geq 0, \forall m \neq n, k, k', \mathbf{a} \in \mathcal{A}$, (or $\frac{\partial f_n^k(\mathbf{a}_{-n})}{\partial a_m^{k'}} \leq 0, \forall m \neq n, k, k', \mathbf{a} \in \mathcal{A}$), if

$$\rho(\bar{\mathbf{T}}^{\max}) < 1, \quad \text{(C5)}$$

then best response dynamics converges linearly to the unique NE of game Γ, for any set of initial conditions belonging to $\mathcal{A}$ with either sequential or parallel updates.

Theorem 6: For $h_n^k(\cdot)$ defined in (20), if

$$\rho(\bar{\mathbf{S}}^{\max}) < 1, \quad \text{(C6)}$$

then best response dynamics converges linearly to the unique NE of game Γ, for any set of initial conditions belonging to $\mathcal{A}$ with either sequential or parallel updates.

Remark 6: (Implications of conditions (C4), (C5), and (C6)) By mean value theorem, we know that the upper bound of additive sum of first derivatives $\sum_{k=1}^K \left| \frac{\partial f_n^k(\mathbf{a}_{-n})}{\partial a_m^{k'}} \right|$ governs the maximum impact that user m's action can make over user n's utility. As a result, Theorem 4, Theorem 5, and Theorem 6 indicate that $\sum_{k=1}^K \left| \frac{\partial f_n^k(\mathbf{a}_{-n})}{\partial a_m^{k'}} \right|$ can be used to develop similar sufficient conditions for the global convergence of best response dynamics. Table II summarizes the connections and differences between all the aforementioned conditions from (C1) to (C6). We can verify that, for the linear function $f_n^k(\cdot)$ that is defined in (11) and studied in Section III-A, $\forall \mathbf{a} \in \mathcal{A}, m \neq n$, it satisfies

$$\frac{\partial f_n^k(\mathbf{a}_{-n})}{\partial a_m^{k'}} = \begin{cases} F_{mn}^k, & \text{if } k' = k \\ 0, & \text{otherwise.} \end{cases} \quad (29)$$

In addition, we can see that, in Example 4, $f_n^k(\cdot)$ is actually a affine function with

$$\frac{\partial f_n^k(\mathbf{P}_{-n})}{\partial P_m^{k'}} = \begin{cases} \gamma(k-k') H_{mn}^{k'}, & \text{if } k' = k \\ 0, & \text{otherwise.} \end{cases} \quad (30)$$

and $\bar{\mathbf{S}}^{\max}$ is reduced to

$$[\bar{\mathbf{S}}^{\max}]_{mn} \triangleq \begin{cases} \max_{k'} \sum_{k=1}^K \gamma(k-k') H_{mn}^{k'}, & \text{if } m \neq n \\ 0, & \text{otherwise.} \end{cases} \quad (31)$$

As an immediate result of Theorem 6, we have the following corollary which specifies a sufficient condition that guarantees the convergence of iterative water-filling algorithm for the asynchronous transmission in multi-carrier systems [20].

Corollary 1: In Example 4, if the matrix $\bar{\mathbf{S}}^{\max}$ defined in (31) satisfies

$$\rho(\bar{\mathbf{S}}^{\max}) < 1, \quad (32)$$

then iterative water-filling algorithm converges linearly to the unique NE of game Γ, for any set of initial conditions belonging to $\mathcal{A}$ and with either sequential or parallel updates.

C. Connections to the Results of Rosen [30], McKenzie [31], Gabay [32], and Facchinei [13]

1) Gradient play and Rosen's DSC conditions: As we mentioned earlier, ACSCG is actually a special subclass of Rosen's concave games. In [30], Rosen proposed a continuous-time gradient projection based iterative algorithm to obtain a pure NE under the assumption of DSC conditions. Here we present a discrete version of the algorithm in [30], named "*gradient play*". Specifically, at stage t, each user first determines the gradient of its own utility function $u_n(\mathbf{a}_n, \mathbf{a}_{-n}^{t-1})$. Then each user updates its action a_n^t using gradient projection according to

$$a_n^{\prime k,t} = a_n^{k,t-1} + \kappa_n \frac{\partial u_n(\mathbf{a}_n, \mathbf{a}_{-n}^{t-1})}{\partial a_n^k}. \quad (33)$$

and

$$\mathbf{a}_n^t = [a_n^{1,t} a_n^{2,t} \cdots a_n^{K,t}] = \left[a_n^{\prime 1,t} a_n^{\prime 2,t} \cdots a_n^{\prime K,t} \right]_{\mathcal{A}_n}^{\|\cdot\|_2}, \quad (34)$$

where κ_n is the stepsize and $[\mathbf{v}]_{\mathcal{A}_n}^{\|\cdot\|_2}$ denotes the projection of the vector $\mathbf{v}$ onto user n's action set $\mathcal{A}_n$ with respect to the Euclidean norm $\|\cdot\|_2$. If κ_n is chosen to be sufficiently small, gradient play approximates the continuous-time gradient projection algorithm. For each nonnegative vector $\kappa = [\kappa_1 \ldots \kappa_N]$, define

$$g(\mathbf{a}, \kappa) = [\kappa_1 \nabla_1 u_1(\mathbf{a}) \ \kappa_2 \nabla_2 u_2(\mathbf{a}) \ \ldots \ \kappa_N \nabla_N u_N(\mathbf{a})]^T. \quad (35)$$

The definition of DSC given in [30] is that, for fixed $\kappa > 0$, if for every $\mathbf{a}^0, \mathbf{a}^1 \in \mathcal{A}$, we have

$$(\mathbf{a}^1 - \mathbf{a}^0)^T g(\mathbf{a}^0, \kappa) + (\mathbf{a}^0 - \mathbf{a}^1)^T g(\mathbf{a}^1, \kappa) > 0. \quad (36)$$

A sufficient condition for DSC provided by Rosen is that the symmetric matrix $G(\mathbf{a}, \kappa) + G^T(\mathbf{a}, \kappa)$ be negative definite for $\mathbf{a} \in \mathcal{A}$, where $G(\mathbf{a}, \kappa)$ is the Jacobian with respect to $\mathbf{a}$ of $g(\mathbf{a}, \kappa)$.

2) McKenzie's dominant diagonal conditions in ACSCG: We notice that Jacobi matrix $G(\mathbf{a}, \kappa)$ with a dominant diagonal is sufficient to guarantee that $G(\mathbf{a}, \kappa) + G^T(\mathbf{a}, \kappa)$ is negative definite. The general definition of matrices with dominant diagonals are proposed by McKenzie in [31]. In particular, a matrix A is said to have a dominant diagonal if there exist numbers $d_n > 0$ such that

$$d_n|a_{nn}| > \sum_{m \neq n} d_m |a_{mn}|. \quad (37)$$

By Theorem 2 in [31], it is immediate that $A + A^T$ is negative definite. As a result, the DSC conditions hold and a pure NE can be found using gradient play for sufficiently small κ_n.

In ACSCG games, it is easy to verify equation (38) in which the elements in $\bar{\mathbf{T}}^{mn}$ are defined as

$$[\bar{\mathbf{T}}^{mn}]_{k'k} = \frac{\partial f_n^k(\mathbf{a}_{-n})}{\partial a_m^{k'}}. \quad (39)$$

$$[\bar{\mathbf{T}}^{\max}]_{mn} \triangleq \begin{cases} \max_{\mathbf{a} \in \mathcal{A}, k'} \sum_{k=1}^{K} \left| \frac{\partial f_n^k(\mathbf{a}_{-n})}{\partial a_m^{k'}} \right|, & \text{if } m \neq n \\ 0, & \text{otherwise.} \end{cases} \quad (40)$$

3) Facchinei's variational approach: In [13] and the references therein, the uniqueness and convergence of NE are proved based on the variational inequality approach. However, the conditions in [13] (e.g. strong monotonicity or P-property) are very generally difficult to check and have no straightforward interpretation for communication networks. The conditions required for convergence in our manuscript are easy to check and have a clear communication meaning (e.g. measure of the strength of the mutual coupling between users).

D. Connections to Linearly Coupled Communication Games

We investigated in [27] the convergence properties in certain communication scenarios, namely *linearly coupled communication games* (LCCG), in which each user has a convex action set $\mathcal{A}_n \subseteq \mathcal{R}_+$ and the utility functions take the form

$$u_n(\mathbf{a}) = a_n^{\beta_n} \cdot (\mu - \sum_{m=1}^{N} \tau_m a_m). \quad (41)$$

It has been used to model the flow control of various classes of traffic in communication networks [34]. In best response dynamics, at stage t, user n chooses its action according to

$$B_n(\mathbf{a}^{t-1}) = \frac{\beta_n(\mu - \sum_{m \in \mathcal{N} \setminus \{n\}} \tau_m a_m^{t-1})}{\tau_n(1 + \beta_n)}. \quad (42)$$

We can see that, LCCG is similar to ACSCG in the sense that the best response iterations at stage t in (14) and (42) both contain the linear combinations of $\mathbf{a}^{t-1}$. However, since $\mathcal{A}_n \subseteq \mathcal{R}$ in LCCG, we can explicitly derive the Jacobian matrix for best response dynamics and determine the exactly locations of all its eigenvalues. Consequently, we are able to develop the necessary and sufficient condition that ensures the spectral radius of the Jacobian matrix to be less than 1 and best response dynamics globally converges. However, in ACSCG, due to the sum-constraint, there exists a non-linear operation $[x]_b^a$ in equation (14). This complicates the analysis of the Jacobian matrix's eigenvalues. Therefore, we usually choose various appropriate matrix norms to bound the spectral radius of the Jacobian matrix and ensure the best response iteration converge under these matrix norms. This approach generally results in various sufficient, but not necessary, conditions.

IV. Solutions with Message Passing

In this section, our objective is to coordinate the users' actions in ACSCG to maximize the overall performance of the system, measured in terms of their total utilities. Specifically, the optimization problem we want to solve is

$$\max_{\mathbf{a} \in \mathcal{A}} \sum_{n=1}^{N} u_n(\mathbf{a}). \quad (43)$$

We will study two distributed algorithms in which the participating users exchange price signals that indicate the "cost" or "benefit" that its action causes to the other users. Allocating network resources via pricing has been well-investigated for convex NUM problems [14], where the original NUM problem can be decomposed into distributively solvable subproblems by setting price for each constraint resources, and each subproblem has to decide the amount of resources to be used depending on the charged price. However, unlike in conventional NUM, pricing mechanisms may not be immediately applicable in ACSCG if utility functions are non-concave. Therefore, we are interested in characterizing the convergence condition of different pricing algorithms in ACSCG.

We know that for any local maximum $\mathbf{a}^*$ of problem (43), there exist unique Lagrange multipliers $\lambda_n, \nu_n^1, \cdots, \nu_n^N$ and $\nu_n'^1, \cdots, \nu_n'^N$ such that the following Karush-Kuhn-Tucker (KKT) conditions hold for all $n \in \mathcal{N}$:

$$\frac{\partial u_n(\mathbf{a}^*)}{\partial a_n^k} + \sum_{m \neq n} \frac{\partial u_m(\mathbf{a}^*)}{\partial a_n^k} = \lambda_n + \nu_n^k - \nu_n'^k, \ \forall n \quad (44)$$

$$\lambda_n \Big(\sum_{k=1}^{K} a_n^{k*} - M_n \Big) = 0, \ \lambda_n \geq 0 \quad (45)$$

$$\nu_n^k (a_n^{k*} - a_{n,k}^{\max}) = 0, \ \nu_n'^k (a_{n,k}^{\min} - a_n^{k*}) = 0, \ \nu_n^k, \nu_n'^k \geq 0. \quad (46)$$

Denote π_{mn}^k user m's marginal fluctuation in utility per unit decrease in user n's action a_n^k within the kth dimension

$$\pi_{mn}^k(a_m^k, \mathbf{a}_{-m}^k) = -\frac{\partial u_m(\mathbf{a})}{\partial a_n^k}, \quad (47)$$

which can be viewed as the cost charged (or compensation paid) to user n for changing user m's utility. Using (47), equation (44) can be rewritten as

$$\frac{\partial u_n(\mathbf{a}^*)}{\partial a_n^k} - \sum_{m \neq n} \pi_{mn}^k (a_m^{k*}, \mathbf{a}_{-m}^{k*}) = \lambda_n + \nu_n^k - \nu_n'^k. \quad (48)$$

If we assume fixed prices $\{\pi_{mn}^k\}$ and action profile $\mathbf{a}_{-n}^k$, condition (48) gives the necessary and sufficient KKT condition of the following problem:

$$\max_{\mathbf{a}_n \in \mathcal{A}_n} \ u_n(\mathbf{a}) - \sum_{k=1}^{K} a_n^k \Big(\sum_{m \neq n} \pi_{mn}^k \Big). \quad (49)$$

$$G^T(\mathbf{a},\kappa) = \begin{bmatrix} \frac{\partial^2 h_1^1}{\partial^2 x} & 0 & \dots & 0 \\ \vdots & \frac{\partial^2 h_1^2}{\partial^2 x} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{\partial^2 h_N^K}{\partial^2 x} \end{bmatrix} \cdot \begin{bmatrix} \kappa_1 I_K & \kappa_2 \bar{\mathbf{T}}^{12} & \dots & \kappa_N \bar{\mathbf{T}}^{1N} \\ \vdots & \kappa_2 I_K & \dots & \kappa_N \bar{\mathbf{T}}^{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \kappa_1 \bar{\mathbf{T}}^{N1} & \kappa_2 \bar{\mathbf{T}}^{N2} & \dots & \kappa_N I_K \end{bmatrix}. \quad (38)$$

At an optimum, a user behaves as if it maximize the differences between its utility minus its payment to the other users in the network due to its impact over the other users' utilities. Different distributed pricing mechanisms can be developed based on the individual objective function in (49) and the convergence conditions may also vary based on the specific action update equation. As we mentioned earlier, since the optimization problem in (43) is not necessarily convex, the pricing algorithms developed for convex NUM, e.g. gradient and subgradient algorithms, can not be directly applied. In the next two subsections, we will make use of two distributed pricing mechanisms and provide the sufficient conditions that guarantee their convergence to feasible operating points that satisfy the KKT conditions. For ease of presentation, similarly as in Section III-A, we temporarily assume $f_n^k(\mathbf{a}_{-n})$ takes the form in (11).

A. Gradient Play

The first distributed pricing algorithm that we consider is the gradient play algorithm. As opposed to equations (33) and (34), the update iterations of gradient play need to be properly redefined in presence of real-time information exchange. Specifically, at stage t, users adopting this algorithm exchange price signals $\{\pi_{mn}^{k,t-1}\}$ using the gradient information at stage $t-1$. Within each iteration, each user first determines the gradient of the objective in (49) based on the price vectors $\{\pi_{mn}^{k,t-1}\}$ and its own utility function $u_n(\mathbf{a}_n, \mathbf{a}_{-n}^{t-1})$. Then each user updates its action a_n^t using gradient projection algorithm according to

$$a_n^{\prime k,t} = a_n^{k,t-1} + \kappa\Big(\frac{\partial u_n(\mathbf{a}_n, \mathbf{a}_{-n}^{t-1})}{\partial a_n^k} - \sum_{m\neq n} \pi_{mn}^{k,t-1}\Big). \quad (50)$$

and

$$\mathbf{a}_n^t = [a_n^{1,t} a_n^{2,t} \cdots a_n^{K,t}] = \Big[a_n^{\prime 1,t} a_n^{\prime 2,t} \cdots a_n^{\prime K,t}\Big]_{\mathcal{A}_n}^{\|\cdot\|_2}. \quad (51)$$

in which the stepsize $\kappa > 0$. The following theorem provides a sufficient condition under which gradient play algorithm will converge monotonically provided that we choose small enough positive constant κ.

Theorem 7: If $\forall n, k, \mathbf{x}, \mathbf{y} \in \mathcal{A}_{-n}$,

$$\inf_x \frac{\partial^2 h_n^k(x)}{\partial^2 x} > -\infty, \text{ and } \Big\|\bigtriangledown g_n^k(\mathbf{x}) - \bigtriangledown g_n^k(\mathbf{y})\Big\| \leq L' \big\|\mathbf{x}-\mathbf{y}\big\|, \quad \text{(C7)}$$

gradient play converges for a small enough stepsize κ.

Proof: It can be proved by showing the gradient of the objective function in (43) is Lipschitz continuous and applying Proposition 3.4 in [26]. See Appendix E for details. ■

Remark 7: (Application of condition (C7)) A sufficient condition that guarantees the convergence of distributed gradient projection algorithm is the Lipschitz continuity of the gradient of the objective function in (43). For example, in the power control problem in multi-channel networks [22], we have $h_n^k(x) = \log_2(\alpha_n^k + H_{nn}^k x)$ and $g_n^k(\mathbf{P}_{-n}) = \log_2(\sigma_n^k + \sum_{m\neq n} H_{mn}^k P_m^k)$. For this configuration, we can immediately verify that condition (C7) is satisfied. Therefore, this gradient play algorithm can be applied. Moreover, as in [22], if we can further ensure that the problem in (43) is convex for some particular utility functions, the gradient play algorithm converges to the unique optimal solution of (43) at which achieving KKT conditions implies global optimality.

B. Jacobi Update

We consider another alternative strategy update mechanism called Jacobi update [28]. In Jacobi update, every user adjusts its action gradually towards the best response strategy. Specifically, the maximizer of problem (49) takes the following form

$$B_n^{\prime k}(\mathbf{a}_{-n}) = \Big[\frac{\partial h_n^k}{\partial x}\Big]^{-1}\big(\lambda_n + \nu_n^k - \nu_n^{\prime k} + \sum_{m\neq n} \pi_{mn}^k\big) - \sum_{m\neq n} F_{mn}^k a_m^k, \quad (52)$$

in which λ_n, ν_n^k, and $\nu_n^{\prime k}$ are the Lagrange multipliers that satisfy complementary slackness in (45) and (46), and π_{mn}^k is defined in (47). In Jacobi update, at stage t, user n chooses its action according to

$$a_n^{k,t} = a_n^{k,t-1} + \kappa\big[B_n^{\prime k}(\mathbf{a}_{-n}^{t-1}) - a_n^{k,t-1}\big], \quad (53)$$

in which the stepsize $\kappa > 0$. The following theorem establishes a sufficient convergence condition for Jacobi update.

Theorem 8: If $\forall n, k, \mathbf{x}, \mathbf{y} \in \mathcal{A}_{-n}$, (C8) is satisfied, then Jacobi update converges if the stepsize κ is sufficiently small.

Proof: It can be proved using the descent lemma and mean value theorem. The details of the proof are provided in Appendix F. ■

Remark 8: (Relation between condition (C8) and the result in [23]) Shi et al. consider the power allocation for multicarrier wireless networks with non-separable utilities. Specifically, $u_n(\cdot)$ takes the form

$$u_n(\mathbf{P}) = r_i\Bigg(\sum_{k=1}^K \log_2\Big(1 + \frac{H_{nn}^k P_n^k}{\sigma_n^k + \sum_{m\neq n} H_{mn}^k P_m^k}\Big)\Bigg), \quad (54)$$

in which $r_i(\cdot)$ is an increasing and strictly concave function. Since the utilities are non-separable, the distributed pricing algorithm proposed in [23], which in fact belongs to Jacobi update, requires only one user to update its action profile at each stage while keeping the remaining users' action fixed. The condition in (C8) gives the convergence condition of the same algorithm in ACSCG. We prove in Theorem 7 that, if the utilities are separable, convergence can still be achieved even if these users update their actions at the same time. Therefore,

$$\inf_x \frac{\partial^2 h_n^k(x)}{\partial^2 x} > -\infty, \ \sup_x \frac{\partial^2 h_n^k(x)}{\partial^2 x} < 0, \ \text{and} \ \left\| \bigtriangledown g_n^k(\mathbf{x}) - \bigtriangledown g_n^k(\mathbf{y}) \right\| \leq L' \|\mathbf{x}-\mathbf{y}\|, \tag{C8}$$

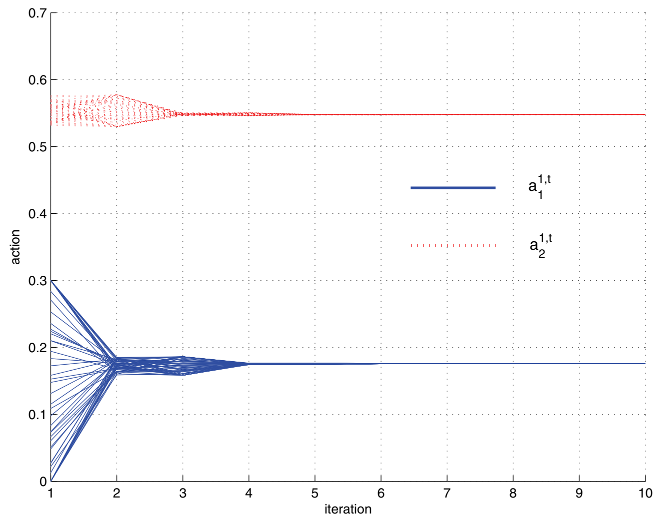

Fig. 2. Actions versus iterations in Example 1.

we do not need an arbitrator to select the only one user that updates its action at each stage.

Remark 9: (Complexity of signaling) The complexity of message exchange measured in terms of the number of price signals to update in (47) is generally of the order of $O(KN^2)$. It is worth mentioning that the amount of signaling can be further reduced to $O(KN)$ in the scenarios where $g_n^k(\cdot)$ are functions of $\sum_{m\neq n} F_{mn}^k a_m^k$. In this case, each user only need to announce one price signal π_n^k for each dimension of its action space:

$$\pi_n^k(a_n^k, \mathbf{a}_{-n}^k) = -\frac{\partial u_n(\mathbf{a})}{\partial\left(\sum_{m\neq n} F_{mn}^k a_m^k\right)} \tag{55}$$

Consequently, π_{mn}^k can be determined based on $\pi_{mn}^k = F_{nm}^k \pi_m^k$, which greatly reduces the overhead of signaling requirement. It is straightforward to check that only $O(KN)$ messages need to be generated and exchanged per iteration in both examples (5) and (6).

Remark 10: (Extension to general cases) As a matter of fact, conditions (C7) and (C8) apply to a broader class of multi-user interaction scenarios, including the general model defined in (4). Specifically, as addressed in Remark 7, the Lipschitz continuity of the gradient of $\sum_{n=1}^N u_n(\mathbf{a})$ is sufficient to guarantee that gradient play with small enough stepsize achieves an operating point at which KKT conditions are satisfied. In addition, we can use the same technique in Appendix F to show the convergence of Jacobi play given that $\sup_x \frac{\partial^2 h_n^k(x)}{\partial^2 x} < 0$, $\forall n, k$, and the gradient of $\sum_{n=1}^N u_n(\mathbf{a})$ is Lipschitz continuous.

V. Numerical Examples

In Section II-C, we present several illustrative examples of ACSCG. This section uses Examples 1 and 3 to illustrate various distributed algorithms addressed in the paper.

We start with Example 1 to verify the proposed convergence conditions of best response dynamics. Even though it is a simple two-user game with $\mathcal{A}_n \subseteq \mathcal{R}^2$, existing results in the literature cannot immediately determine whether or not the best response dynamics in this simple game can globally converge to a NE. Specifically, in Example 1, we have equation (56). Note that in this example, $h_n^k(\cdot)$ belongs to the class of function defined in (20) with $\theta = -0.5$ and $F_{nn}^k = 1$, $\forall n, k$. According the definition of (27) and (28), we have $\bar{\mathbf{S}}^{\max} = \bar{\mathbf{T}}^{\max}$. in which $[\bar{\mathbf{S}}^{\max}]_{12}$ is given in (57). Similarly, we can obtain $[\bar{\mathbf{S}}^{\max}]_{21} = \frac{11}{9} M_2$. Therefore, $\rho(\bar{\mathbf{S}}^{\max}) = \frac{11}{9}\sqrt{M_1 M_2}$. Hence, by condition (C6), we know if $M_1 M_2 < \frac{81}{121}$, the best response dynamics is guaranteed to converge to a unique NE in Example 1. We numerically simulate a scenario with parameters $M_1 = 0.6$ and $M_2 = 1.1$. Note that $M_1 M_2 = 0.66 < \frac{81}{121}$. We generate multiple initial action profiles of $\mathbf{a}_1^0$ and $\mathbf{a}_2^0$, iterate the best response dynamics, and obtain the action sequences $\mathbf{a}_1^t$ and $\mathbf{a}_2^t$. Fig. 2 shows the trajectories of $a_1^{1,t}$ and $a_2^{1,t}$ for different realizations. We can see that, the best response dynamics converges to a unique NE. An interesting phenomenon that can be observed from the analysis above is that, the convergence condition may depend on the maximum constraints M_1 and M_2. This differs from the observation in [21] that the presence of the transmit power and spectral mask constraints does not affect the convergence capability of the iterative water-filling algorithm. This is because when functions $f_n^k(\mathbf{a}_{-n})$ are affine, e.g. in Example 2, 3, and 4, the elements in $\bar{\mathbf{T}}^{\max}$ and $\bar{\mathbf{S}}^{\max}$ are independent of the values of M_1 and M_2. Therefore, (C1)-(C6) are independent of M_n for affine $f_n^k(\mathbf{a}_{-n})$. However, for non-linear $f_n^k(\mathbf{a}_{-n})$, the values of M_1 and M_2 specify the range of users' feasible action sets $\mathcal{A}_1$ and $\mathcal{A}_2$. It will vary $\bar{\mathbf{T}}^{\max}$ and $\bar{\mathbf{S}}^{\max}$ accordingly. In other words, in presence of non-linear coupled $f_n^k(\mathbf{a}_{-n})$, convergence may depend on the players' maximum sum constraints.

Now we consider Example 3, which is the problem of minimizing queueing delays in a Jackson network. In particular, we consider a network with $N = 5$ nodes and $K = 3$ traffic classes. The total routing probability $1 - r_{m0}^k$ that node m will route packets of class k completing service to other nodes is the same for $\forall m \in \mathcal{N}$. We varied the total routing probability $1 - r_{m0}^k$ and generated multiple sets of network parameters in which r_{mn}^k are uniformly distributed for $n = 1, 2, \cdots, N$, μ_n^k are uniformly selected in $[4, 5]$ for $\forall n, k$, and $\psi_n^{\min}$ are uniformly chosen in $[0.6, 1]$ for $n = 1, 2, \cdots, N$.

First of all, we compare the range of validity of the proposed convergence conditions. As we mentioned before, we have $F_{mn}^k = \frac{[(\mathrm{I}-\mathrm{R}^k)^{-1}]_{nm}}{[(\mathrm{I}-\mathrm{R}^k)^{-1}]_{nn}}$ in this example. Note that $(\mathrm{I}-\mathrm{R}^k)^{-1} = \mathrm{I} + \sum_{i=1}^{\infty} (\mathrm{R}^k)^i$ and R^k is a non-negative matrix. Therefore, we can conclude $F_{mn}^k \geq 0, \forall m \neq n, k$. Moreover, since $h_n^k(x) = -\frac{1}{\mu_n^k - v_{nn}^k x}$, we choose to compare conditions (C2) and (C3). In Fig. 3, we plot the probability that conditions (C2) and (C3) are satisfied versus the total routing probability $1 - r_{m0}^k$. From Fig. 3, we can see that the probability of

$$\frac{\partial f_n^1(\mathbf{a}_{-n})}{\partial a_{-n}^1} = \frac{a_{-n}^1}{2}, \frac{\partial f_n^1(\mathbf{a}_{-n})}{\partial a_{-n}^2} = -\frac{2a_{-n}^2}{9}, \frac{\partial f_n^2(\mathbf{a}_{-n})}{\partial a_{-n}^1} = -\frac{2a_{-n}^1}{5}, \frac{\partial f_n^2(\mathbf{a}_{-n})}{\partial a_{-n}^2} = a_{-n}^2. \tag{56}$$

$$[\bar{\mathbf{S}}^{\max}]_{12} = \max\left\{\max_{\mathbf{a}\in\mathcal{A}}\sum_{k=1}^{K}\left|\frac{\partial f_2^k(\mathbf{a}_{-n})}{\partial a_1^1}\right|, \max_{\mathbf{a}\in\mathcal{A}}\sum_{k=1}^{K}\left|\frac{\partial f_2^k(\mathbf{a}_{-n})}{\partial a_1^2}\right|\right\} = \max\left\{\max_{\mathbf{a}_1\in\mathcal{A}_1}\frac{a_1^1}{2}+\frac{2a_1^1}{5}, \max_{\mathbf{a}_1\in\mathcal{A}_1}\frac{2a_1^2}{9}+a_1^2\right\} = \frac{11}{9}M_1. \tag{57}$$

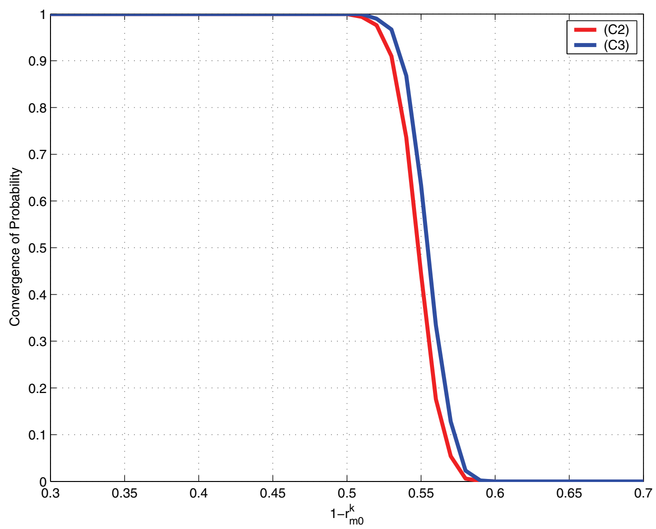

Fig. 3. Probability of (C2) and (C3) versus $1 - r_{m0}^k$ for $\forall m, k$, $N = 5$, $K = 3$.

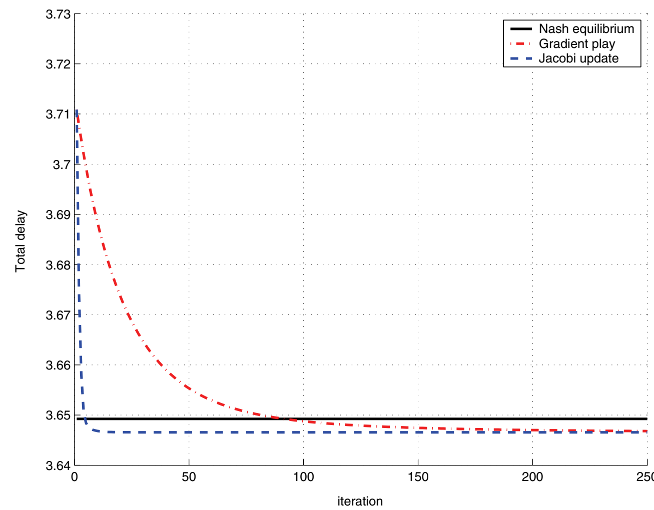

Fig. 5. Illustration of convergence for gradient play and Jacobi update.

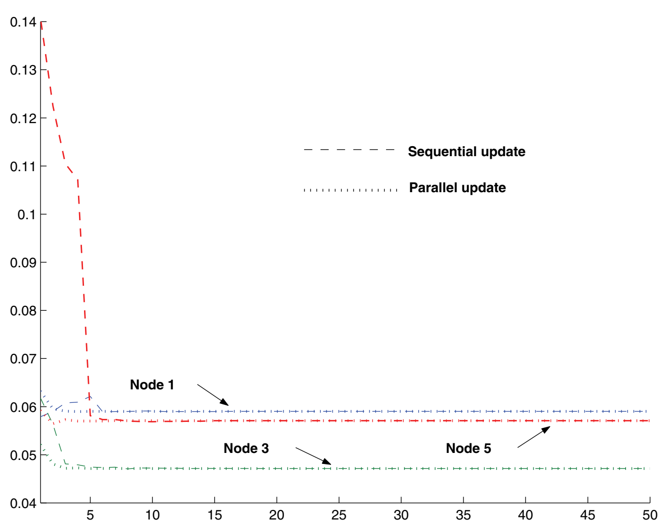

Fig. 4. Delays of nodes versus iterations.

guaranteeing convergence decreases as the routing probability $1 - r_{m0}^k$ increases and condition (C3) shows a similar but slightly broader validity than (C2). Fig. 4 shows three nodes' delay trajectories using both sequential and parallel updates in a certain network realization in which (C2) and (C3) are satisfied. We can see that, the parallel update converges faster than the sequential update.

In Fig. 3, we also note that the probability that (C2) or (C3) is satisfied transits very quickly from the almost certain convergence to the non-convergence guarantee as $1 - r_{m0}^k$ varies from 0.5 to 0.58. Similar observations have been drawn in the multi-channel power control problem [21], where $\theta = -1$ in (20) and the probability that condition (C3) is satisfied exhibits a neat threshold behavior as the ratio between the source-interferer distance and the source-destination distance varies. In Jackson networks, this threshold can be roughly estimated. Define $[\mathbf{S}^k]_{mn} = F_{mn}^k$ for $m \neq n$ and $[\mathbf{S}^k]_{nn} = 0$ for $n \in \mathcal{N}$. If we fix $1 - r_{m0}^k$ for $\forall m, k$, we prove in Appendix G that $\rho(\mathbf{S}^k) \leq \frac{1}{r_{m0}^k} - 1$ for $\forall k$. Therefore, $\rho(\mathbf{S}^k) < 1$ when $r_{m0}^k > 0.5$. We would like to estimate $\rho(\mathbf{T}^{\max})$ and $\rho(\mathbf{S}^{\max})$ based on $\rho(\mathbf{S}^k)$. Note that $\mathbf{T}^{\max}$ defined in (18) is the element-wise maximum over $\mathbf{S}^k$ for $k = 1, 2, \ldots, K$. Since $\mathbf{T}^{\max}$ and $\mathbf{S}^k$ non-negative matrices, we know that $\rho(\mathbf{T}^{\max}) \geq \max_k \rho(\mathbf{S}^k)$. In addition, recall the effect of $\max_{m,n} \zeta_{mn}$ discussed in Remark 5. We can approximate $\rho(\mathbf{S}^{\max})$ defined in (22) using $\rho(\mathbf{S}^{\max}) \approx \max_{m,n} \zeta_{mn} \max_k \rho(\mathbf{S}^k)$. Therefore, we expect that $\rho(\mathbf{T}^{\max})$ and $\rho(\mathbf{S}^{\max})$ exceeds 1 for $r_{m0}^k < 0.5$, which agrees with our observation from Fig. 3. The physical interpretation is that, if the packets exit the network with a probability less than 50% after completing its service, i.e. more than half of the served packets will be routed to other nodes, the strength of the mutual coupling among users becomes too strong and the multi-user interaction in Jackson networks will gradually lose its convergence guarantee.

In addition, we numerically compare two distributed algorithms in which users pass coordination messages in real time, including Jacobi update and gradient play. Fig. 5 shows the delay evolution of both distributed solutions for a particular simulated network in which we set $\kappa = 0.2$. We initialize the system parameters such that $\inf_{n,k} \mu_n^k - \sum_{m=1}^{N} \upsilon_{mn}^k \psi_m^k > 0$

and both conditions (C7) and (C8) are satisfied. We can see that, in spite of the non-concave nature of problem (43), both algorithms cause the total marginal delay to monotonically decrease until it reaches an operating point at which KKT conditions (44)-(46) are satisfied. Using the same stepsize κ, Jacobi play converges more quickly than gradient play in this example. Similar observations are drawn in the other simulated examples, which is because the update directions of these two algorithms are different. Jacobi play algorithm moves directly towards the optimal solution of (49), which is a local approximation of the original optimization program in (43), whereas gradient play algorithm simply updates the actions along the gradient direction of (43).

VI. Conclusion

In this paper, we propose and investigate a game model named additively coupled sum constrained games in which each player is subjected to a sum constraint and its payoff are additively impacted by the remaining users actions. The convergence properties of various generic distributed adjustment algorithms, including best response, gradient play, and Jacobi update, have been addressed. The sufficient conditions obtained in this paper generalize the existing results developed in the multi-channel power control problem and can be extended to other applications that belong to ACSCG.

Appendix A
Proof of Theorem 1

The following lemma is needed to prove Theorem 1.

Lemma 1: Consider any non-decreasing function $p(x)$ and non-increasing function $q(x)$. If there exists a unique x^* such that $p(x^*) = q(x^*)$, and the functions $p(x)$ and $q(x)$ are strictly increasing and strictly decreasing at $x = x^*$ respectively, then $x^* = \arg\min_x\{\max\{p(x), q(x)\}\}$.

Proof of Lemma 1: See Lemma 1 in [20]. ■

Denote $a_n^{k,t}$ as the action of user n in the kth dimension after iteration t. Recall that $\left[h_n^k\right]'(\cdot) > 0$, for $\forall n, k$. Therefore, $\sum_{k=1}^K a_n^{k,t} = M_n$ is satisfied at the end of any iteration t for any user n. Define $[x]^+ = \max\{x, 0\}$ and $[x]^- = \max\{-x, 0\}$. It is straightforward to see that

$$\sum_{k=1}^K [a_n^{k,t} - a_n^{k,t'}]^+ = \sum_{k=1}^K [a_n^{k,t} - a_n^{k,t'}]^-, \forall n, t, t'. \quad (58)$$

We also define

$$p^{n,t}(x) \triangleq \sum_{k=1}^K \left[l_n^k(\mathbf{a}_{-n}^t, x) - a_n^{k,t}\right]^- \quad (59)$$

and

$$q^{n,t}(x) \triangleq \sum_{k=1}^K \left[l_n^k(\mathbf{a}_{-n}^t, x) - a_n^{k,t}\right]^+, \quad (60)$$

in which $l_n^k(\cdot)$ is defined in (14). Since $h_n^k(\cdot)$ is a continuous increasing and strictly concave function, it is clear that $\left\{\frac{\partial h_n^k}{\partial x}\right\}^{-1}(\cdot)$ is a continuous decreasing function. If $p^{n,t}(\lambda_n^{t+1}) \neq 0$ (i.e. it has not converged), $p^{n,t}(x)$ ($q^{n,t}(x)$, respectively) is non-decreasing (non-increasing) in x, and strictly increasing (strictly decreasing) at $x = \lambda_n^{t+1}$. From (58) it is always true that $p^{n,t}(\lambda_n^{t+1}) = q^{n,t}(\lambda_n^{t+1})$. We first prove the convergence of the parallel update case in (17). For $\forall n$, we have equations and equalities (61)-(68). where (61) and (68) follows from (58), (62) follows from the definition of $p^{n,t}$ and $q^{n,t}$ in (59) and (60), (63) is due to Lemma 1 in which $x = \lambda_n^t$, (64) follows from the definition of $p^{n,t}$ and $q^{n,t}$, the expression of $a_n^{k,t}$ in (17), and the fact that $\left[[x]_a^b - [y]_a^b\right]^+ \leq [x-y]^+$ and $\left[[x]_a^b - [y]_a^b\right]^- \leq [x-y]^-$, (65) is due to the fact that $[x+y]^+ \leq [x]^+ + [y]^+$ and $[x+y]^- \leq [x]^- + [y]^-$, (67) follows by using $[\sum_k x_k y_k]^+ \leq \sum_k |x_k||y_k| = \sum_k |x_k|([y_k]^+ + [y_k]^-) \leq \max_k |x_k| \sum_k([y_k]^+ + [y_k]^-)$. For user n, we define that $e_n^t = \left[a_m^{k,t} - a_m^{k,t-1}\right]^+$. Inequality (68) can be written as $e_n^{t+1} \leq \sum_{m\neq n}[\mathbf{T}^{\max}]_{mn} e_m^t$ in which $\mathbf{T}^{\max}$ is defined in (18).

Since $\mathbf{T}^{\max}$ is a nonnegative matrix, by the Perron-Frobenius Theorem [26], there exists a positive vector $\bar{\mathbf{w}} = [\bar{w}_1 \dots \bar{w}_N]$ such that

$$\|\mathbf{T}^{\max}\|_{\infty,\text{mat}}^{\bar{\mathbf{w}}} = \rho(\mathbf{T}^{\max}), \quad (69)$$

where $\|\cdot\|_{\infty,\text{mat}}^{\bar{\mathbf{w}}}$ is the weighted maximum matrix norm defined as

$$\|\mathbf{A}\|_{\infty,\text{mat}}^{\bar{\mathbf{w}}} \triangleq \max_{i=1,2,\cdots,N} \frac{1}{\bar{w}_i} \sum_{j=1}^N [\mathbf{A}]_{ij} \bar{w}_j, \quad \mathbf{A} \in \mathcal{R}^{N\times N}. \quad (70)$$

Define the vectors $\mathbf{e}^{t+1} \triangleq [e_1^{t+1}, e_2^{t+1}, \dots, e_N^{t+1}]^T$ and $\mathbf{e}^t \triangleq [e_1^t, e_2^t, \dots, e_N^t]^T$. The set of inequalities in (68) can be expressed in the vector form as $\mathbf{0} \leq \mathbf{e}^{t+1} \leq \mathbf{T}^{\max}\mathbf{e}^t$. By choosing the vector $\bar{\mathbf{w}}$ that satisfies $\|\mathbf{T}^{\max}\|_{\infty,\text{mat}}^{\bar{\mathbf{w}}} = \rho(\mathbf{T}^{\max})$ and applying the infinity norm $\|\cdot\|_\infty^{\bar{\mathbf{w}}}$, we obtain the following

$$\|\mathbf{e}^{t+1}\|_\infty^{\bar{\mathbf{w}}} \leq 2\|\mathbf{T}^{\max}\mathbf{e}^t\|_\infty^{\bar{\mathbf{w}}} \leq 2\|\mathbf{T}^{\max}\|_{\infty,\text{mat}}^{\bar{\mathbf{w}}} \|\mathbf{e}^t\|_\infty^{\bar{\mathbf{w}}}, \quad (71)$$

Finally, based on (68) and (71), it follows that

$$\begin{aligned}\max_{n\in\mathcal{N}} \frac{e_n^{t+1}}{\bar{w}_n} &= \|\mathbf{e}^{t+1}\|_\infty^{\bar{\mathbf{w}}} \leq 2\|\mathbf{T}^{\max}\|_{\infty,\text{mat}}^{\bar{\mathbf{w}}} \|\mathbf{e}^t\|_\infty^{\bar{\mathbf{w}}} \\ &\leq 2\|\mathbf{T}^{\max}\|_{\infty,\text{mat}}^{\bar{\mathbf{w}}} \cdot \max_{n\in\mathcal{N}} \frac{e_n^t}{\bar{w}_n} = 2\rho(\mathbf{T}^{\max}) \cdot \max_{n\in\mathcal{N}} \frac{e_n^t}{\bar{w}_n} \quad (72)\end{aligned}$$

Therefore, if $\|\mathbf{T}^{\max}\|_{\infty,\text{mat}}^{\bar{\mathbf{w}}} = \rho(\mathbf{T}^{\max}) < \frac{1}{2}$, the best response dynamics in (17) is a contraction with the modulus $\|\mathbf{T}^{\max}\|_{\infty,\text{mat}}^{\bar{\mathbf{w}}}$ with respect to the norm $\max_{n\in\mathcal{N}} \frac{\|\cdot\|_2^{\mathbf{w}_n}}{\bar{w}_n}$. We can conclude that, the best response dynamics has a unique fixed point $\mathbf{a}^*$ and, given any initial value $\mathbf{a}^0$, the update sequence $\{\mathbf{a}^t\}$ converges to the fixed point $\mathbf{a}^*$.

In the sequential update case, the convergence result can be established by using the proposition 1.4 in [26]. The key step is to obtain

$$\max_{n\in\mathcal{N}} \frac{e_n^{t+1}}{\bar{w}_n} \leq 2\rho(\mathbf{T}^{\max}) \cdot \max\left\{\max_{j<n} \frac{e_j^{t+1}}{\bar{w}_j}, \max_{j\geq n} \frac{e_j^t}{\bar{w}_j}\right\}. \quad (73)$$

A simple induction on n yields

$$\max_{n\in\mathcal{N}} \frac{e_n^{t+1}}{\bar{w}_n} \leq 2\rho(\mathbf{T}^{\max}) \cdot \max_{n\in\mathcal{N}} \frac{e_n^t}{\bar{w}_n} \quad (74)$$

for all n. Therefore, inequality (68) also holds for the sequential update and the contraction iteration globally converges to a unique equilibrium. ■

$$\sum_{k=1}^{K}[a_n^{k,t+1}-a_n^{k,t}]^+ = \max\Big\{\sum_{k=1}^{K}[a_n^{k,t+1}-a_n^{k,t}]^+, \sum_{k=1}^{K}[a_n^{k,t+1}-a_n^{k,t}]^-\Big\} \tag{61}$$

$$= \max\{p^{n,t}(\lambda_n^{t+1}), q^{n,t}(\lambda_n^{t+1})\} \tag{62}$$

$$\leq \max\{p^{n,t}(\lambda_n^{t}), q^{n,t}(\lambda_n^{t})\} \tag{63}$$

$$\leq \max\Big\{\sum_{k=1}^{K}\Big[\sum_{m\neq n}F_{mn}^k(a_m^{k,t}-a_m^{k,t-1})\Big]^+, \sum_{k=1}^{K}\Big[\sum_{m\neq n}F_{mn}^k(a_m^{k,t}-a_m^{k,t-1})\Big]^-\Big\} \tag{64}$$

$$\leq \max\Big\{\sum_{k=1}^{K}\sum_{m\neq n}\Big[F_{mn}^k(a_m^{k,t}-a_m^{k,t-1})\Big]^+, \sum_{k=1}^{K}\sum_{m\neq n}\Big[F_{mn}^k(a_m^{k,t}-a_m^{k,t-1})\Big]^-\Big\} \tag{65}$$

$$= \max\Big\{\sum_{m\neq n}\sum_{k=1}^{K}\Big[F_{mn}^k(a_m^{k,t}-a_m^{k,t-1})\Big]^+, \sum_{m\neq n}\sum_{k=1}^{K}\Big[F_{mn}^k(a_m^{k,t}-a_m^{k,t-1})\Big]^-\Big\} \tag{66}$$

$$\leq \sum_{m\neq n}\max_k|F_{mn}^k|\cdot\Big\{\sum_{k=1}^{K}\Big[a_m^{k,t}-a_m^{k,t-1}\Big]^+ + \sum_{k=1}^{K}\Big[a_m^{k,t}-a_m^{k,t-1}\Big]^-\Big\} \tag{67}$$

$$= \sum_{m\neq n}2\max_k|F_{mn}^k|\cdot\sum_{k=1}^{K}\Big[a_m^{k,t}-a_m^{k,t-1}\Big]^+, \tag{68}$$

APPENDIX B
PROOF OF THEOREM 2

If $F_{mn}^k \geq 0, \forall m\neq n, k$, the inequalities after (66) become

$$\max\Big\{\sum_{m\neq n}\sum_{k=1}^{K}\Big[F_{mn}^k(a_m^{k,t}-a_m^{k,t-1})\Big]^+, \sum_{m\neq n}\sum_{k=1}^{K}\Big[F_{mn}^k(a_m^{k,t}-a_m^{k,t-1})\Big]^-\Big\} \tag{75}$$

$$\leq \sum_{m\neq n}\max_k F_{mn}^k\cdot\max\Big\{\sum_{k=1}^{K}\Big[a_m^{k,t}-a_m^{k,t-1}\Big]^+, \sum_{k=1}^{K}\Big[a_m^{k,t}-a_m^{k,t-1}\Big]^-\Big\} \tag{76}$$

$$= \sum_{m\neq n}\max_k F_{mn}^k\cdot\sum_{k=1}^{K}\Big[a_m^{k,t}-a_m^{k,t-1}\Big]^+. \tag{77}$$

Similarly, for $F_{mn}^k \leq 0, \forall m\neq n, k$, we have

$$\max\Big\{\sum_{m\neq n}\sum_{k=1}^{K}\Big[F_{mn}^k(a_m^{k,t}-a_m^{k,t-1})\Big]^+, \sum_{m\neq n}\sum_{k=1}^{K}\Big[F_{mn}^k(a_m^{k,t}-a_m^{k,t-1})\Big]^-\Big\} \tag{78}$$

$$\leq \sum_{m\neq n}\max_k\{-F_{mn}^k\}\cdot\max\Big\{\sum_{k=1}^{K}\Big[a_m^{k,t}-a_m^{k,t-1}\Big]^+, \sum_{k=1}^{K}\Big[a_m^{k,t}-a_m^{k,t-1}\Big]^-\Big\} \tag{79}$$

$$= \sum_{m\neq n}\max_k\{-F_{mn}^k\}\cdot\sum_{k=1}^{K}\Big[a_m^{k,t}-a_m^{k,t-1}\Big]^+. \tag{80}$$

Therefore, if $F_{mn}^k \geq 0, \forall m\neq n, k$ or $F_{mn}^k \leq 0, \forall m\neq n, k$, given (C2), the sequence $\{\mathbf{a}_n^t\}$ contracts with the modulus $\rho(\mathbf{T}^{\max}) < 1$ under the norm $\max_{n\in\mathcal{N}}\frac{\sum_k[x_n^k]^+}{\bar{w}_n}$ and the convergence follows readily. ■

APPENDIX C
PROOF OF THEOREM 3

Let $\|\cdot\|_2^{\mathbf{w}}$ denote the weighted Euclidean norm with weights $\mathbf{w}=[w_1\dots w_K]^T$, i.e. $\|\mathbf{x}\|_2^{\mathbf{w}} \triangleq (\sum_i w_i|x_i|^2)^{1/2}$ [25]. Define the simplex (81) in which $\sum_k x_k^{\max} \geq 1$. The following lemma is needed to prove Theorem 3.

Lemma 2: The projection with respect to the weighted Euclidean norm with weights $\mathbf{w}$, of the K-dimensional real vector $-\mathbf{x}_0 \triangleq -[x_{0,1},\dots,x_{0,K}]^T$ onto the simplex $\mathcal{S}$ defined in (81), denoted by $[-\mathbf{x}_0]_{\mathcal{S}}^{\mathbf{w}}$, is the optimal solution to the following convex optimization problem:

$$[-\mathbf{x}_0]_{\mathcal{S}}^{\mathbf{w}} \triangleq \arg\min_{\mathbf{x}\in\mathcal{S}}\big\|\mathbf{x}-(-\mathbf{x}_0)\big\|_2^{\mathbf{w}} \tag{82}$$

and takes the following form:

$$x_k^* = \Big[\frac{\lambda}{w_k}-x_{0,k}\Big]_{x_k^{\min}}^{x_k^{\max}}, k=1,\dots,K \tag{83}$$

where $\lambda > 0$ is chosen in order to satisfy the constraint $\frac{1}{K}\sum_{k=1}^{K}x_k^* = 1$.

Proof of Lemma 2: See Corollary 2 in [21]. ■

For $h_n^k(\cdot)$ defined in (20), user n updates its action according to (84) and λ^* is chosen to satisfy $\sum_{k=1}^{K}a_n^{*k} = M_n$. Define the vector update operator as $[\mathrm{BR}(\mathbf{a}_{-n})]_k \triangleq a_n^{*k}$ and the coupling vector as

$$[\mathbf{C}_n(\mathbf{a}_{-n})]_k \triangleq \frac{\alpha_n^k}{F_{nn}^k}+\sum_{m\neq n}F_{mn}^k a_m^k \tag{85}$$

with $k\in\{1,\dots,K\}$. We also define

$$\mathbf{F}'_{mn} \triangleq \mathrm{diag}\Big(F_{mn}^1, F_{mn}^2, \dots, F_{mn}^K\Big) \tag{86}$$

$$\mathcal{S} \triangleq \left\{ \mathbf{x} \in \mathcal{R}^K : \frac{1}{K}\sum_{k=1}^{K} x_k = 1, x_k^{\min} \leq x_k \leq x_k^{\max},\ \forall k = 1, 2, \ldots, K \right\}. \tag{81}$$

$$a_n^{*k} = l_n^k(\mathbf{a}_{-n}, \lambda^*) = \Big[\big(\frac{1}{F_{nn}^k}\big)^{1+\frac{1}{\theta}} \cdot (\lambda^*)^{\frac{1}{\theta}} - \frac{\alpha_n^k}{F_{nn}^k} - \sum_{m \neq n} F_{mn}^k a_m^k \Big]_{a_{n,k}^{\min}}^{a_{n,k}^{\max}}. \tag{84}$$

and

$$\boldsymbol{\alpha}_n' \triangleq \Big[\frac{\alpha_n^1}{F_{nn}^1}, \frac{\alpha_n^2}{F_{nn}^2}, \ldots, \frac{\alpha_n^K}{F_{nn}^K} \Big]^T. \tag{87}$$

Therefore, the coupling vector can be alternatively rewritten as

$$\mathbf{C}_n(\mathbf{a}_{-n}) = \boldsymbol{\alpha}_n' + \sum_{m \neq n} \mathbf{F}_{mn}' \mathbf{a}_m. \tag{88}$$

Define a weight matrix $\mathbf{W} = [\mathbf{w}_1 \ldots \mathbf{w}_N]$ in which the element $[\mathbf{W}]_{kn}$ is chosen according to

$$[\mathbf{W}]_{kn} = [\mathbf{w}_n]_k = \frac{(F_{nn}^k)^{1+\frac{1}{\alpha}}}{\sum_{k=1}^K (F_{nn}^k)^{1+\frac{1}{\alpha}}}. \tag{89}$$

By Lemma 2, we know that the vector update operator $\mathrm{BR}_n(\mathbf{a}_{-n})$ in (21) can be interpreted as the projection of the coupling vector $-\mathbf{C}_n(\mathbf{a}_{-n})$ onto user n's action set $\mathcal{A}_n$ with respect to $\|\cdot\|_2^{\mathbf{w}_n}$, i.e.

$$\mathrm{BR}_n(\mathbf{a}_{-n}) = \big[-\mathbf{C}_n(\mathbf{a}_{-n}) \big]_{\mathcal{A}_n}^{\mathbf{w}_n}. \tag{90}$$

Given any $\mathbf{a}^{(1)}, \mathbf{a}^{(2)} \in \mathcal{A}$, we define respectively, for each user n, the weighted Euclidean distances between these two vectors and their projected vectors using (90) as $e_n = \big\| \mathbf{a}_n^{(2)} - \mathbf{a}_n^{(1)} \big\|_2^{\mathbf{w}_n}$ and $e_{\mathrm{BR}_n} = \big\| \mathrm{BR}_n(\mathbf{a}_{-n}^{(1)}) - \mathrm{BR}_n(\mathbf{a}_{-n}^{(2)}) \big\|_2^{\mathbf{w}_n}$. Again, we first prove the convergence of the parallel update case in (17). We have $\forall n \in \mathcal{N}$,

$$\begin{aligned}
e_{\mathrm{BR}_n} &= \Big\| \big[-\mathbf{C}_n(\mathbf{a}_{-n}^{(1)}) \big]_{\mathcal{A}_n}^{\mathbf{w}_n} - \big[-\mathbf{C}_n(\mathbf{a}_{-n}^{(2)}) \big]_{\mathcal{A}_n}^{\mathbf{w}_n} \Big\|_2^{\mathbf{w}_n} \\
&\leq \Big\| \mathbf{C}_n(\mathbf{a}_{-n}^{(2)}) - \mathbf{C}_n(\mathbf{a}_{-n}^{(1)}) \Big\|_2^{\mathbf{w}_n} && (91) \\
&= \Big\| \sum_{m \neq n} \mathbf{F}_{mn}' \mathbf{a}_m^{(2)} - \sum_{m \neq n} \mathbf{F}_{mn}' \mathbf{a}_m^{(1)} \Big\|_2^{\mathbf{w}_n} \\
&= \Big\| \sum_{m \neq n} \mathbf{F}_{mn}' \big(\mathbf{a}_m^{(2)} - \mathbf{a}_m^{(1)} \big) \Big\|_2^{\mathbf{w}_n} && (92) \\
&\leq \sum_{m \neq n} \Big\| \mathbf{F}_{mn}' \big(\mathbf{a}_m^{(2)} - \mathbf{a}_m^{(1)} \big) \Big\|_2^{\mathbf{w}_n} \\
&= \sum_{m \neq n} \sqrt{ \sum_{k=1}^K [\mathbf{w}_n]_k \big([\mathbf{F}_{mn}']_{kk} \big)^2 \Big(a_m^{(2)k} - a_m^{(1)k} \Big)^2 } && (93) \\
&= \sum_{m \neq n} \sqrt{ \sum_{k=1}^K [\mathbf{w}_m]_k \Big([\mathbf{F}_{mn}']_{kk} \frac{[\mathbf{w}_n]_k}{[\mathbf{w}_m]_k} \Big)^2 \Big(a_m^{(2)k} - a_m^{(1)k} \Big)^2 } && (94) \\
&\leq \sum_{m \neq n} \max_k \Big(\big| [\mathbf{F}_{mn}']_{kk} \big| \cdot \frac{[\mathbf{w}_n]_k}{[\mathbf{w}_m]_k} \Big) \sqrt{ \sum_{k=1}^K [\mathbf{w}_m]_k \Big(a_m^{(2)k} - a_m^{(1)k} \Big)^2 } && (95) \\
&= \sum_{m \neq n} \max_k \Big(\big| [\mathbf{F}_{mn}']_{kk} \big| \cdot \frac{[\mathbf{w}_n]_k}{[\mathbf{w}_m]_k} \Big) \big\| \mathbf{a}_m^{(2)} - \mathbf{a}_m^{(1)} \big\|_2^{\mathbf{w}_m} && (96) \\
&= \sum_{m \neq n} [\mathbf{S}^{\max}]_{mn} e_m, && (97)
\end{aligned}$$

where (91) follows from the non-expansion property of the projector $[\cdot]_{\mathcal{A}_n}^{\mathbf{w}_n}$ in the norm $\|\cdot\|_2^{\mathbf{w}_n}$ (See Proposition 3.2(c) in [26]), (93) follows from the triangle inequality [25], and $\mathbf{S}^{\max}$ in (97) is defined according to (22).

The rest of the proof is similar as the proof after equation (68) in Appendix A. Details are omitted due to space limitations. ■

APPENDIX D
PROOF OF THEOREM 4

The beginning part of the proof is the same as the proof of Theorem 1. For any user n with general $f_n^k(\cdot)$, the inequalities after (62) become (98 - 103) where (98) follows from the definition of $p^{n,t}$ and $q^{n,t}$ and the expression of $a_n^{k,t}$ and $B_n^k(\mathbf{a}_{-n}, \lambda)$ in (17) and (26), (99) follows from the mean value theorem for vector-valued functions with $\xi^t = \alpha \mathbf{a}^t + (1-\alpha)\mathbf{a}^{t-1}$ and $\alpha \in [0,1]$. By (C4), it is straightforward to show that the iteration is a contraction by following the same arguments in Appendix A. The rest of the proof is omitted. ■

APPENDIX E
PROOF OF THEOREM 6

The gradient play algorithm in (49) is in fact a gradient projection algorithm with constant stepsize κ. In order to establish its convergence, we first need to prove that the gradient of the objective in (43) is Lipschitz continuous, with a Lipschitz constant given by $L > 0$, i.e.

$$\Big\| \triangledown \big(\sum_{n=1}^N u_n(\mathbf{x}) \big) - \triangledown \big(\sum_{n=1}^N u_n(\mathbf{y}) \big) \Big\| \leq L \big\| \mathbf{x} - \mathbf{y} \big\|,\ \forall \mathbf{x}, \mathbf{y} \in \mathcal{A}. \tag{104}$$

It is known that it has the property of Lipschitz continuity if it has a Hessian bounded in the Euclidean norm.

The Hessian matrix $\mathbf{H}$ of $\sum_{n=1}^N u_n(\mathbf{a})$ can be decomposed into two matrices: $\mathbf{H} = \mathbf{H}_1 + \mathbf{H}_2$, in which the elements of matrix $\mathbf{H}_1$ are defined in (105) with $F_{nn}^k = 1$ and the elements of matrix $\mathbf{H}_2$ are

$$-\frac{\partial^2 \Big[\sum_{n=1}^N \sum_{k=1}^K g_n^k(\mathbf{a}_{-n}) \Big]}{\partial a_m^k \partial a_l^j} = -\sum_{n=1}^N \sum_{k=1}^K \frac{\partial^2 g_n^k(\mathbf{a}_{-n})}{\partial a_m^k \partial a_l^j}. \tag{106}$$

Recall that $g_n^k(\cdot)$ is Lipschitz continuous and it satisfies

$$\Big\| \triangledown g_n^k(\mathbf{x}) - \triangledown g_n^k(\mathbf{y}) \Big\| \leq L' \big\| \mathbf{x} - \mathbf{y} \big\|,\ \forall n, k, \mathbf{x}, \mathbf{y} \in \mathcal{A}_{-n}.$$

Consequently, we have $\|\mathbf{H}_2\|_2 \leq NKL'$. As a result, we can estimate the Lipschitz constant L using the inequalities in (107). We can choose the RHS of (107) as the Lipschitz

$$\sum_{k=1}^{K}[a_n^{k,t+1}-a_n^{k,t}]^+ \leq \max\{p^{n,t}(\lambda_n^t), q^{n,t}(\lambda_n^t)\}$$

$$= \max\left\{\sum_{k=1}^{K}\left[f_n^k(\mathbf{a}_{-n}^t)-f_n^k(\mathbf{a}_{-n}^{t-1})\right]^+, \sum_{k=1}^{K}\left[f_n^k(\mathbf{a}_{-n}^t)-f_n^k(\mathbf{a}_{-n}^{t-1})\right]^-\right\} \tag{98}$$

$$= \max\left\{\sum_{k=1}^{K}\Big[\sum_{m\neq n}\sum_{k'=1}^{K}\frac{\partial f_n^k}{\partial a_m^{k'}}(\xi_{-n}^t)(a_m^{k',t}-a_m^{k',t-1})\Big]^+, \sum_{k=1}^{K}\Big[\sum_{m\neq n}\sum_{k'=1}^{K}\frac{\partial f_n^k}{\partial a_m^{k'}}(\xi_{-n}^t)(a_m^{k',t}-a_m^{k',t-1})\Big]^-\right\} \tag{99}$$

$$\leq \max\left\{\sum_{k=1}^{K}\sum_{m\neq n}\sum_{k'=1}^{K}\Big[\frac{\partial f_n^k}{\partial a_m^{k'}}(\xi_{-n}^t)(a_m^{k',t}-a_m^{k',t-1})\Big]^+, \sum_{k=1}^{K}\sum_{m\neq n}\sum_{k'=1}^{K}\Big[\frac{\partial f_n^k}{\partial a_m^{k'}}(\xi_{-n}^t)(a_m^{k',t}-a_m^{k',t-1})\Big]^-\right\} \tag{100}$$

$$= \max\left\{\sum_{m\neq n}\sum_{k'=1}^{K}\sum_{k=1}^{K}\Big[\frac{\partial f_n^k}{\partial a_m^{k'}}(\xi_{-n}^t)(a_m^{k',t}-a_m^{k',t-1})\Big]^+, \sum_{m\neq n}\sum_{k'=1}^{K}\sum_{k=1}^{K}\Big[\frac{\partial f_n^k}{\partial a_m^{k'}}(\xi_{-n}^t)(a_m^{k',t}-a_m^{k',t-1})\Big]^-\right\} \tag{101}$$

$$\leq \sum_{m\neq n}\left\{\max_{k'}\sum_{k=1}^{K}\Big|\frac{\partial f_n^k}{\partial a_m^{k'}}(\xi_{-n}^t)\Big|\right\}\cdot\left\{\sum_{k'=1}^{K}\Big[(a_m^{k',t}-a_m^{k',t-1})\Big]^+ + \sum_{k'=1}^{K}\Big[(a_m^{k',t}-a_m^{k',t-1})\Big]^-\right\} \tag{102}$$

$$= \sum_{m\neq n}2\cdot\left\{\max_{k'}\sum_{k=1}^{K}\Big|\frac{\partial f_n^k}{\partial a_m^{k'}}(\xi_{-n}^t)\Big|\right\}\cdot\sum_{k'=1}^{K}\Big[(a_m^{k',t}-a_m^{k',t-1})\Big]^+ \tag{103}$$

$$\frac{\partial^2\Big[\sum_{n=1}^{N}\sum_{k=1}^{K}h_n^k\big(a_n^k+\sum_{m\neq n}F_{mn}^k a_m^k\big)\Big]}{\partial a_m^k\partial a_l^j} = \begin{cases}\sum_{n=1}^{N}\frac{\partial^2 h_n^k}{\partial^2 x}\big(a_n^k+\sum_{m\neq n}F_{mn}^k a_m^k\big)F_{mn}^k F_{ln}^k, & \text{if } k=j\\ 0, & \text{otherwise.}\end{cases} \tag{105}$$

$$\|\mathbf{H}\|_2 \leq \|\mathbf{H}_1\|_2+\|\mathbf{H}_2\|_2 \leq \sqrt{\|\mathbf{H}_1\|_1\|\mathbf{H}_1\|_\infty}+NKL' \leq \sup_{x,n,k}\Big|\frac{\partial^2 h_n^k}{\partial^2 x}\Big|\cdot\max_{k,l}\sum_{m=1}^{N}\sum_{n=1}^{N}|F_{mn}^k F_{ln}^k|+NKL'. \tag{107}$$

constant L . By Proposition 3.4 in [26], we know that if $0<\kappa<2/L$, the sequence $\mathbf{a}^t$ generated by the gradient projection algorithm in (50) and (51) converges to a limiting point at which the KKT conditions in (44)-(46) are satisfied. ■

APPENDIX F
PROOF OF THEOREM 7

We know from the proof of Theorem 6 that, under Condition (C7), $\sum_{n=1}^{N}u_n(\mathbf{a})$ is Lipschitz continuous and the inequality in (104) holds. Recall that $\sum_{n=1}^{N}u_n(\mathbf{x})$ is continuously differentiable. Therefore, by the descent lemma [26], we have $\forall\mathbf{x},\mathbf{y}\in\mathcal{A}$

$$\sum_{n=1}^{N}u_n(\mathbf{x}) \geq \sum_{n=1}^{N}u_n(\mathbf{y})+(\mathbf{x}-\mathbf{y})^T\cdot\bigtriangledown\big(\sum_{n=1}^{N}u_n(\mathbf{y})\big)-\frac{L}{2}\|\mathbf{x}-\mathbf{y}\|_2^2. \tag{108}$$

Therefore, in order to prove $\sum_{n=1}^{N}u_n(\mathbf{a}^t)\geq\sum_{n=1}^{N}u_n(\mathbf{a}^{t-1})$, we only need to show that

$$(\mathbf{a}^t-\mathbf{a}^{t-1})^T\cdot\bigtriangledown\big(\sum_{n=1}^{N}u_n(\mathbf{a}^{t-1})\big)\geq\frac{L}{2}\big\|\mathbf{a}^t-\mathbf{a}^{t-1}\big\|_2^2 \tag{109}$$

for sufficiently small κ. Substituting (53) into (109), we can see that it is equivalent to

$$\sum_{n=1}^{N}\sum_{k=1}^{K}\big(B_n'^k(\mathbf{a}_{-n}^{t-1})-a_n^{k,t-1}\big)\cdot\frac{\partial\sum_{n=1}^{N}u_n(\mathbf{a}^{t-1})}{\partial a_n^{k,t-1}} \geq \kappa\cdot\frac{L}{2}\cdot\sum_{n=1}^{N}\sum_{k=1}^{K}\big(B_n'^k(\mathbf{a}_{-n}^{t-1})-a_n^{k,t-1}\big)^2. \tag{110}$$

By equation (52), we have

$$B_n^{'k}(\mathbf{a}_{-n}^{t-1}) - a_n^{k,t-1} = \Big[\frac{\partial h_n^k}{\partial x}\Big]^{-1}\big(\lambda_n + \nu_n^k - \nu_n^{'k} + \sum_{m\neq n}\pi_{mn}^{k,t-1}\big) - \sum_{m\neq n}^{N} F_{mn}^k a_m^{k,t-1} - a_n^{k,t-1} \quad (111)$$

and

$$\frac{\partial \sum_{n=1}^N u_n(\mathbf{a}^{t-1})}{\partial a_n^{k,t-1}} = \frac{\partial h_n^k}{\partial x}\big(a_n^{k,t-1} + \sum_{m\neq n}^{N} F_{mn}^k a_m^{k,t-1}\big) - \sum_{m\neq n}\pi_{mn}^{k,t-1}. \quad (112)$$

By the mean value theorem, there exists $\xi_n^k \in \mathcal{R}$ such that

$$\begin{aligned}
&\frac{\partial h_n^k}{\partial x}\big(a_n^{k,t-1} + \sum_{m\neq n}^{N} F_{mn}^k a_m^{k,t-1}\big) - \sum_{m\neq n}\pi_{mn}^{k,t-1} \\
&= \frac{\partial h_n^k}{\partial x}\big(a_n^{k,t-1} + \sum_{m\neq n}^{N} F_{mn}^k a_m^{k,t-1}\big) \\
&- \Big(\lambda_n + \nu_n^k - \nu_n^{'k} + \sum_{m\neq n}\pi_{mn}^{k,t-1}\Big) + \Big(\lambda_n + \nu_n^k - \nu_n^{'k}\Big) \\
&= \frac{\partial^2 h_n^k}{\partial^2 x}(\xi_n^k)\cdot\Bigg\{a_n^{k,t-1} + \sum_{m\neq n}^{N} F_{mn}^k a_m^{k,t-1} \\
&- \Big[\frac{\partial h_n^k}{\partial x}\Big]^{-1}\Big(\lambda_n + \nu_n^k - \nu_n^{'k} + \sum_{m\neq n}\pi_{mn}^{k,t-1}\Big)\Bigg\} + \lambda_n + \nu_n^k - \nu_n^{'k}.
\end{aligned}$$

Multiplying (111) and (112) leads to

$$\begin{aligned}
&\sum_{n=1}^{N}\sum_{k=1}^{K}\big(B_n^{'k}(\mathbf{a}_{-n}^{t-1}) - a_n^{k,t-1}\big)\cdot\frac{\partial \sum_{n=1}^N u_n(\mathbf{a}^{t-1})}{\partial a_n^{k,t-1}} \\
&= -\sum_{n=1}^{N}\sum_{k=1}^{K}\frac{\partial^2 h_n^k}{\partial^2 x}(\xi_n^k)\cdot\Bigg\{\Big[\frac{\partial h_n^k}{\partial x}\Big]^{-1}\Big(\lambda_n + \nu_n^k - \nu_n^{'k} \\
&+ \sum_{m\neq n}\pi_{mn}^{k,t-1}\Big) - a_n^{k,t-1} - \sum_{m\neq n}^{N} F_{mn}^k a_m^{k,t-1}\Bigg\}^2 \\
&+ \sum_{n=1}^{N}\sum_{k=1}^{K}\big(B_n^{'k}(\mathbf{a}_{-n}^{t-1}) - a_n^{k,t-1}\big)\cdot(\lambda_n + \nu_n^k - \nu_n^{'k}) \quad (113)
\end{aligned}$$

In the following, we differentiate two cases in which the Lagrange multipliers $\lambda_n, \nu_n^k, \nu_n^{'k}$ take different values.

First of all, if $\lambda_n = \nu_n^k = \nu_n^{'k} = 0$ for all k, n, equation (113) can be simplified as

$$\begin{aligned}
&\sum_{n=1}^{N}\sum_{k=1}^{K}\big(B_n^{'k}(\mathbf{a}_{-n}^{t-1}) - a_n^{k,t-1}\big)\cdot\frac{\partial \sum_{n=1}^N u_n(\mathbf{a}^{t-1})}{\partial a_n^{k,t-1}} \\
&= -\sum_{n=1}^{N}\sum_{k=1}^{K}\frac{\partial^2 h_n^k}{\partial^2 x}(\xi_n^k)\cdot\Bigg\{\Big[\frac{\partial h_n^k}{\partial x}\Big]^{-1}\Big(\lambda_n + \nu_n^k - \nu_n^{'k} \\
&+ \sum_{m\neq n}\pi_{mn}^{k,t-1}\Big) - a_n^{k,t-1} - \sum_{m\neq n}^{N} F_{mn}^k a_m^{k,t-1}\Bigg\}^2. \quad (114)
\end{aligned}$$

On the other hand, if $\lambda_n > 0$, $\nu_n^k > 0$, or $\nu_n^{'k} > 0$ for some k, n. Due to complementary slackness in (45) and (46), We know that

$$\begin{aligned}
&\lambda_n > 0 \Rightarrow \sum_{k=1}^{K} B_n^{'k}(\mathbf{a}_{-n}^{t-1}) = M_n \geq \sum_{k=1}^{K} a_n^{k,t-1}, \\
&\nu_n^k > 0 \Rightarrow B_n^{'k}(\mathbf{a}_{-n}^{t-1}) = a_{n,k}^{\max} \geq a_n^{k,t-1}, \\
&\nu_n^{'k} > 0, \Rightarrow B_n^{'k}(\mathbf{a}_{-n}^{t-1}) = a_{n,k}^{\min} \leq a_n^{k,t-1}.
\end{aligned}$$

As a result, the last term in (113) satisfy

$$\begin{aligned}
&\sum_{n=1}^{N}\sum_{k=1}^{K}\big(B_n^{'k}(\mathbf{a}_{-n}^{t-1}) - a_n^{k,t-1}\big)\cdot(\lambda_n + \nu_n^k - \nu_n^{'k}) \\
&= \sum_{n=1}^{N}\lambda_n\sum_{k=1}^{K}\big(B_n^{'k}(\mathbf{a}_{-n}^{t-1}) - a_n^{k,t-1}\big) \\
&+ \sum_{n=1}^{N}\sum_{k=1}^{K}\nu_n^k\big(B_n^{'k}(\mathbf{a}_{-n}^{t-1}) - a_n^{k,t-1}\big) \\
&+ \sum_{n=1}^{N}\sum_{k=1}^{K}\nu_n^{'k}\big(a_n^{k,t-1} - B_n^{'k}(\mathbf{a}_{-n}^{t-1})\big) \geq 0. \quad (115)
\end{aligned}$$

Therefore, in both cases, the following inequality holds

$$\begin{aligned}
&\sum_{n=1}^{N}\sum_{k=1}^{K}\big(B_n^{'k}(\mathbf{a}_{-n}^{t-1}) - a_n^{k,t-1}\big)\cdot\frac{\partial \sum_{n=1}^N u_n(\mathbf{a}^{t-1})}{\partial a_n^{k,t-1}} \\
&\geq -\sum_{n=1}^{N}\sum_{k=1}^{K}\sup_x\frac{\partial^2 h_n^k(x)}{\partial^2 x}\cdot\Bigg\{\Big[\frac{\partial h_n^k}{\partial x}\Big]^{-1}\Big(\lambda_n + \nu_n^k - \nu_n^{'k} \\
&+ \sum_{m\neq n}\pi_{mn}^{k,t-1}\Big) - a_n^{k,t-1} - \sum_{m\neq n}^{N} F_{mn}^k a_m^{k,t-1}\Bigg\}^2 \\
&= -\sum_{n=1}^{N}\sum_{k=1}^{K}\sup_x\frac{\partial^2 h_n^k(x)}{\partial^2 x}\cdot\big(B_n^{'k}(\mathbf{a}_{-n}^{t-1}) - a_n^{k,t-1}\big)^2. \quad (116)
\end{aligned}$$

Finally, we can conclude that the inequality in (110) holds for $\kappa \leq \frac{2}{L}\cdot(-\max_{n,k}\sup_x\frac{\partial^2 h_n^k(x)}{\partial^2 x})$. Recall that Jacobi update requires $\kappa \in (0, 1]$. The stepsize κ can be eventually chosen as $0 < \kappa \leq \min\{\frac{2}{L}\cdot(-\max_{n,k}\sup_x\frac{\partial^2 h_n^k(x)}{\partial^2 x}), 1\}$ ■

APPENDIX G
UPPER BOUND OF $\rho(\mathbf{S}^k)$

Denote $\mathbf{1}^T = [1\ 1 \cdots 1]^T$. If we fix $1 - r_{m0}^k$ for $\forall m, k$, we have $\mathbf{1}^T\mathrm{R}^k = (1 - r_{m0}^k)\mathbf{1}^T$. Note that $\Upsilon^k = (\mathrm{I} - \mathrm{R}^k)^{-1} = \mathrm{I} + \sum_{i=1}^{\infty}(\mathrm{R}^k)^i$. We have $\mathbf{1}^T\Upsilon^k = \mathbf{1}^T\big(\mathrm{I} + \sum_{i=1}^{\infty}(\mathrm{R}^k)^i\big) = \mathbf{1}^T + (1 - r_{m0}^k)\mathbf{1}^T\Upsilon^k$ and $\mathbf{1}^T\Upsilon^k = \frac{1}{1-r_{m0}^k}\mathbf{1}^T$. Therefore, $|\Upsilon^k|_1 = \frac{1}{r_{m0}^k}$. Since $F_{mn}^k = \frac{[\Upsilon^k]_{nm}}{[\Upsilon^k]_{nn}}$ and $\Upsilon^k = \mathrm{I} + \sum_{i=1}^{\infty}(\mathrm{R}^k)^i$, we know $[\Upsilon^k]_{nn} \geq 1$ for $\forall n$. Denote a diagonal matrix $\mathrm{diag}(\Upsilon^k)$ with the entries of Υ^k on the diagonal. Recall that $[\mathbf{S}^k]_{mn} = F_{mn}^k$ for $m \neq n$, and $[\mathbf{S}^k]_{nn} = 0$ for $n \in \mathcal{N}$. We can conclude that $\rho(\mathbf{S}^k) \leq |\mathbf{S}^k|_\infty \leq |(\Upsilon^k)^T - \mathrm{diag}(\Upsilon^k)|_\infty \leq |(\Upsilon^k)^T|_\infty - 1 = |\Upsilon^k|_1 - 1 = \frac{1}{r_{m0}^k} - 1$.

REFERENCES

[1] E. Altman, T. Boulogne, R. El-Azouzi, T. Jimenez, and L. Wynter, "A survey on networking games in telecommunications," *Comput. Operation Res.*, vol. 33, pp. 286–311, Feb. 2006.

[2] A. MacKenzie and S. Wicker, "Game theory and the design of self-configuring, adaptive wireless networks," *IEEE Commun. Mag.*, vol. 39, pp. 126–131, Nov. 2001.

[3] M. Felegyhazi and J. P. Hubaux, "Game theory in wireless networks: a tutorial," EPFL Technical Report, LCA-REPORT-2006-002, Feb. 2006.

[4] N. Nisan, T. Roughgarden, E. Tardos, and V. Vazirani, *Algorithmic Game Theory*. Cambridge University Press, 2007.

[5] E. A. Jorswieck, E. G. Larsson, M. Luise, and H. V. Poor, "Game theory in signal processing and communications," *IEEE Signal Process. Mag.*, vol. 26, no. 5, Sep. 2009.

[6] D. Fudenberg and J. Tirole, *Game Theory*. MIT Press, 1991.

[7] D. M. Topkis, *Supermodularity and Complementarity*. Princeton University Press, 1998.

[8] D. Yao, "S-modular games with queueing applications," *Queueing Syst.*, vol. 21, pp. 449–475, 1995.

[9] E. Altman and Z. Altman, "S-modular games and power control in wireless networks," *IEEE Trans. Automatic Control*, vol. 48, no. 5, pp. 839–842, May 2003.

[10] R. Rosenthal, "A class of games possessing pure-strategy Nash equilibria," *International J. Game Theory*, vol. 2, pp. 65–67, 1973.

[11] D. Monderer and L. S. Shapley, "Potential games," *Games Econ. Behav.*, vol. 14, no. 1, pp. 124–143, May 1996.

[12] G. Scutari, S. Barbarossa, and D. P. Palomar, "Potential games: a framework for vector power control problems with coupled constraints," *Proc. 2006 IEEE ICASSP*.

[13] F. Facchinei and J.-S. Pang, "Nash equilibria: the variational approach," in *Convex Optimization Signal Process. Commun.*, D. P. Palomar and Y. C. Eldar, editors. Cambridge University Press, 2009, ch. 12.

[14] M. Chiang, S. H. Low, A. R. Calderbank, and J. C. Doyle, "Layering as optimization decomposition," *Proc. IEEE*, vol. 95, pp. 255–312. Jan. 2007.

[15] D. P. Palomar and M. Chiang, "A tutorial on decomposition methods for network utility maximization," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1439–1451, 2006.

[16] S. Lasaulce, M. Debbah, and E. Altman, "Methodologies for analyzing equilibria in wireless games," *IEEE Signal Process. Mag.*, vol. 26, no. 5, pp. 41–52, Sep. 2009.

[17] H. Chen and D. Yao, *Fundamentals of Queueing Networks: Performance, Asymptotics, and Optimization*. Springer, 2001.

[18] W. Yu, G. Ginis, and J. Cioffi, "Distributed multiuser power control for digital subscriber lines," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 5, pp. 1105–1115, June 2002.

[19] S. T. Chung, J. L. Seung, J. Kim, and J. Cioffi, "A game-theoretic approach to power allocation in frequency-selective Gaussian interference channels," in *Proc. 2003 IEEE Int. Symp. Inf. Theory*, p. 316.

[20] R. Cendrillon, J. Huang, M. Chiang, and M. Moonen, "Autonomous spectrum balancing for digital subscriber lines," *IEEE Trans. Signal Process.*, vol. 55, no. 8, pp. 4241–4257, Aug. 2007.

[21] G. Scutari, D. P. Palomar, and S. Barbarossa, "Optimal linear precoding strategies for wideband noncooperative systems based on game theory—part II: algorithms," *IEEE Trans. Signal Process.*, vol. 56, no. 3, pp. 1250–1267, Mar. 2008.

[22] J. Huang, R. Berry, and M. L. Honig, "Distributed interference compensation for wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 5, pp. 1074–1084, May 2006.

[23] C. Shi, R. Berry, and M. Honig, "Distributed interference pricing for OFDM wireless networks with non-seperable utilities," in *Proc. 2008 Conf. Inf. Sciences Syst.*, pp. 755–760.

[24] J. Mo and J. Walrand, "Fair end-to-end window-based congestion control," *IEEE Trans. Netw.*, vol. 8, no. 5, pp. 556–567, Oct. 2000.

[25] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1985.

[26] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation*. Prentice Hall, 1997.

[27] Y. Su and M. van der Schaar, "Linearly coupled communication games," UCLA Technical Report, Aug. 2009.

[28] R. La and V. Anantharam, "Utility based rate control in the Internet for elastic traffic," *IEEE/ACM Trans. Netw.*, vol. 10, no. 2, pp. 271–286, Apr. 2002.

[29] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[30] J. Rosen, "Existence and uniqueness of equilibrium points for concave n-person games," *Econometrica*, vol. 33, pp. 520–534, July 1965.

[31] L. McKenzie, "Matrices with dominant diagonals and economic theory," *Mathematical Methods in the Social Sciences*, Stanford University Press, pp 47–62, 1960

[32] D. Gabay and H. Moulin, "On the uniqueness and stability of Nash equilibria in noncooperative games," *Applied Stochastic Control Econometrics Management Science*. North-Holland, pp. 271–293, 1980.

[33] F. Kelly, A. K. Maulloo, and D. K. H. Tan, "Rate control in communication networks: shadow prices, proportional fairness and stability," *J. Operational Research Society*, vol. 49, pp. 237–252, 1998.

[34] Z. Zhang and C. Douligeris, "Convergence of synchronous and asynchronous greedy algorithm in a multiclass telecommunications environment," *IEEE Trans. Commun.*, vol. 40, pp. 1277–1281, 1992.

[35] J. Bulow, J. Geanakoplos, and P. Klemperer, "Multimarket oligopoly: strategic substitutes and strategic complements," *J. Political Economy*, vol. 93, pp. 488–511, 1985.

Yi Su (S'08) He received the B.E. and M.E. degrees from Tsinghua University, Beijing, China, in 2004 and 2006, respectively, both in electrical engineering. He received the Ph.D. degree in the Department of Electrical Engineering at the University of California, Los Angeles in 2010. He is now with Qualcomm.

Mihaela van der Schaar (F'10) received the Ph.D. degree from Eindhoven University of Technology, The Netherlands, in 2001. She is Chancellors Professor of Electrical Engineering at the University of California, Los Angeles. Her research interests include multimedia networking, communication, processing, and systems, multimedia stream mining, dynamic multi-user networks and system designs, online learning, network economics, and game theory.