Performance Optimization of Distributed Primal-Dual Algorithms over Wireless Networks

Zhaohui Yang*, Mingzhe Chen†‡, Kai-Kit Wong§, Walid Saad¶, H. Vincent Poor†, and Shuguang Cui‡
*Department of Engineering, King's College London, WC2R 2LS, UK.
†Electrical Engineering Department, Princeton University, NJ, 08544, USA.
‡Shenzhen Research Institute of Big Data and School of Science and Engineering, the Chinese University of Hong Kong, Shenzhen, 518172, China.
§Department of Electronic and Electrical Engineering, University College London, London, UK.
¶Wireless@VT, Bradley Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA, 24060, USA.
Emails: yang.zhaohui@kcl.ac.uk, mingzhec@princeton.edu, kai-kit.wong@ucl.ac.uk, walids@vt.edu, poor@princeton.edu, robert.cui@gmail.com.

***Abstract*—In this paper, the implementation of a distributed primal-dual algorithm over realistic wireless networks is investigated. In the considered model, the users and one base station (BS) cooperatively perform a distributed primal-dual algorithm for controlling and optimizing wireless networks. In particular, each user must locally update the primal and dual variables and send the updated primal variables to the BS. The BS aggregates the received primal variables and broadcasts the aggregated variables to all users. Since all of the primal and dual variables as well as aggregated variables are transmitted over wireless links, the imperfect wireless links will affect the solution achieved by the distributed primal-dual algorithm. Therefore, it is necessary to study how wireless factors such as transmission errors affect the implementation of the distributed primal-dual algorithm and how to optimize wireless network performance to improve the solution achieved by the distributed primal-dual algorithm. To address these challenges, the convergence rate of the primal-dual algorithm is first derived in a closed form while considering the impact of wireless factors such as data transmission errors. Based on the derived convergence rate, the optimal transmit power and resource block allocation schemes are designed to minimize the gap between the target solution and the solution achieved by the distributed primal-dual algorithm. Simulation results show that the proposed distributed primal-dual algorithm can reduce the gap between the target and obtained solution by up to 52% compared to the distributed primal-dual algorithm without considering imperfect wireless transmission.**

***Index Terms*—Dual method, convergence rate, resource allocation.**

I. Introduction

Recently, the security and privacy concerns as well as the availability of abundant data and computation resources in wireless networks are pushing the deployment of optimization algorithms towards the network edge [1]. This has led to a significant interest in distributed optimization methods. In distributed optimization, each node can compute on its own data and sends the results to its neighbours or a center node. Distributed optimization has many applications, such as user selection optimization, resource allocation optimization, trajectory optimization, and distributed machine learning algorithm design [2]–[4].

Distributed optimization algorithms fall within two main classes: distributed primal algorithms [5]–[8] and distributed primal-dual algorithms [9]–[13]. In [5], the authors proposed fast distributed gradient algorithms to minimize the sum of individual cost functions. The decentralized gradient descent method was proposed in [8], where all agents collaborate with their neighbors through information exchange. Compared to the distributed primal algorithm, it was shown that a distributed primal-dual algorithm converges rapidly [9]. The distributed alternating direction method of multipliers (ADMM) was proposed in [9] for solving separable optimization problems. For distributed optimization with global inequality constraints, the authors in [10] studied deterministic and stochastic primal-dual sub-gradient algorithms. To reduce the communication cost of a decentralized algorithm, the work in [11] proposed a communication-censored ADMM. A variant ADMM algorithm was proposed in [12] to reduce the communication overhead. To further reduce the communication overhead, the authors in [13] investigated the use of coding techniques for a stochastic incremental distributed primal-dual algorithm. However, most of these existing works [9]–[13] do not consider the effect of imperfect wireless transmission of primal and dual variables on the implementation of distributed primal-dual algorithm over realistic wireless networks. For example, due to the limited transmit power of each user, the transmitted primal and dual variables may contain errors thus affecting the convergence of the distributed primal-dual algorithms and the solution achieved by the distributed primal-dual algorithms.

The main contribution of this paper is a novel framework that enables the implementation of a primal-dual algorithm over a realistic wireless network. In particular, we consider a realistic wireless network that consists of multiple users and one base station (BS) which cooperatively perform a distributed primal-dual algorithm for controlling and optimizing network performance. Here, each user must locally update the primal and dual variables and send the primal variables to the BS, which aggregates the received primal

This work was supported in part by the U.S. National Science Foundation under Grant CCF-1908308, CNS-1814477, by the National Key R&D Program of China with grant No. 2018YFB1800800, by the Key Area R&D Program of Guangdong Province with grant No. 2018B030338001, by Shenzhen Outstanding Talents Training Fund, and by Guangdong Research Project No. 2017ZT07X152.

and broadcasts the aggregated variables to all users. Since all of the primal and dual variables as well as the aggregated variables are transmitted over wireless links, the imperfect wireless links will affect the solution achieved by the distributed primal-dual algorithm. Therefore, it is necessary to study how wireless factors such as transmission errors affect the implementation of the distributed primal-dual algorithm and how to optimize wireless network performance to improve the solution achieved by the distributed primal-dual algorithm. To address these challenges, we first derive the convergence rate of the primal-dual algorithm while considering the impact of wireless factors such as data transmission error. Given the derived convergence rate, we design an optimal transmit power and resource block allocation scheme so as to enable the distributed primal-dual algorithm to find an optimal solution. Simulation results show that the proposed distributed primal-dual algorithm can reduce the gap between the target and the obtained solution by up to 52% compared to the distributed primal-dual algorithm without considering imperfect wireless transmission.

II. System Model and Problem Formulation

Consider a network in which a set $\mathcal{N}$ of N users and one BS jointly implement a distributed primal-dual algorithm. Each user n has a local dataset $\mathcal{D}_n$. Due to data privacy issue, only user n can access dataset $\mathcal{D}_n$.

A. Primal-Dual Model

The users and the BS use the distributed primal-dual algorithm for solving the following optimization problem [11]:

$$\min_{\boldsymbol{x}} \quad f(\boldsymbol{x}) \triangleq \frac{1}{N}\sum_{n=1}^{N} f_n(\boldsymbol{x}, \mathcal{D}_n) \tag{1}$$

$$\text{s.t.} \quad g_m(\boldsymbol{x}) \leq 0, \quad \forall m \in \mathcal{M}, \tag{1a}$$

where $f_n(\boldsymbol{x}, \mathcal{D}_n)$ and $g_m(\boldsymbol{x})$ are convex functions, $\mathcal{M} = \{1, \cdots, M\}$, and M is the number of constraints. For simplicity, we use $f_n(\boldsymbol{x})$ to represent $f_n(\boldsymbol{x}, \mathcal{D}_n)$ in the following.

Using the distributed primal-dual algorithm, the Lagrange function of problem (1) can be given by

$$\begin{aligned}\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) &= \frac{1}{N}\sum_{n=1}^{N} f_n(\boldsymbol{x}) + \sum_{m=1}^{M} \lambda_m g_m(\boldsymbol{x}) \\ &= \frac{1}{N}\sum_{n=1}^{N}\left(f_n(\boldsymbol{x}) + \sum_{m=1}^{M} \lambda_m g_m(\boldsymbol{x})\right),\end{aligned} \tag{2}$$

where $\boldsymbol{\lambda} = [\lambda_1, \cdots, \lambda_M]^T$ is the Lagrange multiplier associated with constraint (1a). For each user n, we define the local Lagrange function as

$$\mathcal{L}_n(\boldsymbol{x}, \boldsymbol{\lambda}) = f_n(\boldsymbol{x}) + \boldsymbol{\lambda}^T \boldsymbol{g}(\boldsymbol{x}), \tag{3}$$

where $\boldsymbol{g}(\boldsymbol{x}) \triangleq [g_1(\boldsymbol{x}), \cdots, g_M(\boldsymbol{x})]^T$. The sub-gradients of local Lagrange function can be given by

$$\nabla_{\boldsymbol{x}} \mathcal{L}_n(\boldsymbol{x}, \boldsymbol{\lambda}) = \nabla f_n(\boldsymbol{x}) + \boldsymbol{\lambda}^T \nabla \boldsymbol{g}(\boldsymbol{x}), \tag{4}$$

and

$$\nabla_{\boldsymbol{\lambda}} \mathcal{L}_n(\boldsymbol{x}, \boldsymbol{\lambda}) = \boldsymbol{g}(\boldsymbol{x}). \tag{5}$$

Algorithm 1 Distributed Primal-Dual Algorithm

1: Initialize primal variable $\boldsymbol{x}(0) = \mathbf{0}$ and dual variable $\boldsymbol{\lambda}(0) = \mathbf{0}$.
2: **for** $t = 0, 1, \cdots, T$
3: **parallel for** user $n \in \mathcal{N}$
4: Update the dual and primal variables:

$$\boldsymbol{\lambda}(t+1) = \boldsymbol{\lambda}(t) + \alpha(t)\boldsymbol{g}(\boldsymbol{x}(t)), \tag{7}$$

$$\boldsymbol{y}_n(t+1) = \boldsymbol{x}(t) - \alpha(t)\nabla_{\boldsymbol{x}}\mathcal{L}_n(\boldsymbol{x}(t), \boldsymbol{\lambda}(t)). \tag{8}$$

5: Each user sends $\boldsymbol{y}_i(t)$ to the BS.
6: **end for**
7: The BS computes

$$\boldsymbol{x}(t+1) = \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{y}_n(t+1) \tag{9}$$

and broadcasts the value to all users.
8: Set $t = t + 1$.
9: **end for**
10: Output weighted average value of the primal variable

$$\hat{\boldsymbol{x}}(T) = \frac{\sum_{t=0}^{T-1} \alpha(t)\boldsymbol{x}(t)}{\sum_{t=0}^{T-1} \alpha(t)}. \tag{10}$$

Based on the definition of the local Lagrange function, the distributed primal-dual algorithm is proposed to solve the following maximin problem [10]:

$$\max_{\boldsymbol{\lambda}} \min_{\boldsymbol{x}} \quad \frac{1}{N}\sum_{n=1}^{N} \mathcal{L}_n(\boldsymbol{x}, \boldsymbol{\lambda}). \tag{6}$$

The distributed primal-dual algorithm used to solve problem (6) is given in Algorithm 1. In Algorithm 1, each user updates the dual variable $\boldsymbol{\lambda}(t+1)$ and obtains a copy of the primal variable $\boldsymbol{y}_n(t+1)$. Note that $\alpha(t)$ is a dynamic step size for the sub-gradient descend procedure. The BS aggregates the obtained copies of primal variables from all users and broadcasts the aggregated vector $\boldsymbol{x}$ to all users. After a sufficient number of iterations, such as T iterations, each user can obtain the primal variable solution as in (10).

B. Wireless Communication Model

For the uplink transmission, orthogonal frequency division multiple access (OFDMA) technique is applied and each user can occupy only one resource block (RB). Assume that the total number of RBs is N. Let $a_{ln} \in \{0, 1\}$ denote the RB association index, i.e., $a_{ln} = 1$ implies that RB l is assigned to user n and $a_{ln} = 0$ otherwise. Since each user can occupy only one RB and each RB should be occupied by only one user, we have

$$\sum_{l=1}^{N} a_{ln} = 1, \sum_{n=1}^{N} a_{ln} = 1. \tag{11}$$

When user n is assigned with RB l, the uplink transmission rate of user n is

$$r_{ln} = B\log_2\left(1 + \frac{p_n \beta_l d_n^{-\zeta} o_n}{I_l + BN_0}\right), \tag{12}$$

where B is the bandwidth of each RB, p_n is the transmission power of user n, β_l is the reference channel gain between the user and the BS on RB l at the reference distance 1 m, d_n is the distance between user n and the BS, ζ is a pathloss factor, and $o_n \sim \exp(1)$ is the small scale fading.

Due to the randomness of wireless communication channel, the user may transmit data with errors. For user n with RB l, the error rate is defined as

$$q_{ln} = \mathbb{P}(r_{ln} < R), \tag{13}$$

where R is the minimum rate for transmitting the updated primal variables to the BS. To calculate the value of q_{ln}, we have the following lemma.

Lemma 1. The data error rate of user n with RB l is

$$q_{ln} = 1 - \exp\left(-\frac{D_{ln}}{p_n}\right), \tag{14}$$

where $D_{ln} = \frac{(2^{R/B}-1)(I_l+BN_0)}{\beta_l d_n^{-\zeta}}$.

Proof: Based on (12) and (13), we have

$$\begin{aligned} q_{ln} &= \mathbb{P}(r_{ln} < R) \\ &= \mathbb{P}\left(o_n < \frac{(2^{R/B}-1)(I_l+BN_0)}{p_n \beta_l d_n^{-\zeta}}\right) \\ &= 1 - \exp\left(-\frac{(2^{R/B}-1)(I_l+BN_0)}{p_n \beta_l d_n^{-\zeta}}\right), \end{aligned} \tag{15}$$

where the last equality follows from $o_n \sim \exp(1)$. ■

Since user n can occupy any one RB, the data error rate of user n is

$$q_n = \sum_{l=1}^{N} a_{ln} q_{ln}. \tag{16}$$

In the considered system, if the received primal variable $\boldsymbol{y}_n$ from user n contains errors, the BS will not use it for the update of the aggregated primal variables. Let $C_n(t) \in \{0, 1\}$ indicate that whether user n transmits primal variable $\boldsymbol{y}_n$ in time t contains error or not. In particular, $C_n(t) = 1$ shows that $\boldsymbol{y}_n$ received by the BS does not contain any data error; otherwise, we have $C_n(t) = 0$. The BS computes the aggregated primal variable as[1]

$$\boldsymbol{x}(t+1) = \frac{\sum_{n=1}^{N} C_n(t)\boldsymbol{y}_n(t+1)}{\sum_{n=1}^{N} C_n(t)}, \tag{17}$$

where

$$C_n(t) = \begin{cases} 1, & \text{with probability } 1 - q_n \\ 0, & \text{with probability } q_n \end{cases}. \tag{18}$$

C. Problem Formulation

We aim to jointly optimize the RB allocation and power control for all users to minimize the gap of the solution achieved by the distributed primal-dual algorithm and the optimal solution that the distributed primal-dual algorithm targets to achieve, which is given as

$$\min_{\boldsymbol{A},\boldsymbol{p}} \quad \mathbb{E}(f(\hat{\boldsymbol{x}}(T)) - f(\boldsymbol{x}^*)) \tag{19}$$

$$\text{s.t.} \quad \sum_{l=1}^{N} a_{ln} = 1, \quad \forall l \in \mathcal{N}, \tag{19a}$$

$$\sum_{n=1}^{N} a_{ln} = 1, \quad \forall n \in \mathcal{N}, \tag{19b}$$

$$\sum_{n=1}^{N} p_n \le P_{\max}, \tag{19c}$$

$$a_{ln} \in \{0, 1\}, \quad \forall l, n \in \mathcal{N}, \tag{19d}$$

$$0 \le p_n \le P_n, \quad \forall n \in \mathcal{N}, \tag{19e}$$

where $\boldsymbol{A} = \{a_{ln}\}_{N \times N}$, $\boldsymbol{p} = [p_1, \cdots, p_N]^T$, $\mathbb{E}(f(\hat{\boldsymbol{x}}(T)) - f(\boldsymbol{x}^*))$ denotes the gap of the solution $\boldsymbol{x}(T)$ achieved by the distributed primal-dual algorithm with T iterations and the optimal solution $\boldsymbol{x}^*$ that the distributed primal-dual algorithm targets to achieve, $P_{\max}$ is the maximum total transmit power of all users, and P_n is the maximum transmit power of user n. Constraints (19a) and (19b) indicate that each user can occupy only one RB and each RB can be assigned with only one user. Constraint (19c) shows that the sum transmit power of all users cannot exceed a given value, which guarantees that the energy consumption of the whole system is limited.

III. Convergence Analysis and Resource Allocation

A. Convergence Analysis

To solve problem (19), we first analyze the convergence of Algorithm 1. To analyze the convergence rate of Algorithm 1, we make the following three assumptions:

Assumption 1. Compact Feasible Set: The feasible set of primal variable $\boldsymbol{x}$ satisfying (1a) is non-empty, compact, and convex. Denote R as the smallest radius of the ℓ_2 ball with original center that contains the feasible set, i.e., $\|\boldsymbol{x}\| \le R$ for all $\boldsymbol{x}$ satisfying (1a). Furthermore, this feasible set is known by all users.

Assumption 2. Slater Condition: There exists a solution $\boldsymbol{x}$ such that $g_m(\boldsymbol{x}) < 0, \forall m \in \mathcal{M}$.

Assumption 2 indicates that the primal problem in (1) and the dual problem (6) have the same optimal objective value, and the optimal dual variable $\boldsymbol{\lambda}^*$ has a finite value. Denote S as the finite maximum value for $\lambda_m(t)$, i.e., $\lambda_m(t) < S$.

Assumption 3. Lipschitz Continuous: Both functions $f_n(\boldsymbol{x})$ and $g_m(\boldsymbol{x})$ are convex on the feasible set, and the first-order derivative of functions $f_n(\boldsymbol{x})$ and $g_m(\boldsymbol{x})$ are bounded by L, i.e.,

$$\nabla f_n(\boldsymbol{x}) \le L, \nabla g_m(\boldsymbol{x}) \le L, \quad \forall n \in \mathcal{N}, m \in \mathcal{M}, \tag{20}$$

where $L < \infty$ is a constant.

Based on the above assumptions, the convergence of Algorithm 1 is shown in the following theorem.

[1]Note that the denominator in (17) is zero only for the case that $C_n(t) = 0$ for all n with probability $\Pi_{n=1}^{N} q_n$. Since the probability $\Pi_{n=1}^{N} q_n$ approaches zero when the number of users is large, we ignore the case that $C_n(t) = 0$ for all n.

Theorem 1. If the BS and the users implement Algorithm 1 over T iterations, the upper bound of $\mathbb{E}(f(\hat{\boldsymbol{x}}(T)) - f(\boldsymbol{x}^*))$ can be given by

$$\mathbb{E}(f(\hat{\boldsymbol{x}}(T)) - f(\boldsymbol{x}^*)) \leq \frac{R^2 + \sum_{n=1}^{N} d_1(1-q_n)}{d_2(1-q_0)} \tag{21}$$

where $d_1 = \sum_{t=0}^{T-1}(L + LMS + ML^2R^2)\alpha(t)^2$, $d_2 = 2N\sum_{t=0}^{T-1}\alpha(t)$ and $q_0 = \max_{n\in\mathcal{N}} q_n$.

Proof: Seen in Appendix A. ■

Theorem 1 provides an upper bound of the gap between $f(\hat{\boldsymbol{x}}(T))$ and $f(\boldsymbol{x}^*)$. If we let the step size $\alpha(t)$ (for example $\alpha(t) = 1/t$) satisfy $\sum_{t=0}^{\infty}\alpha(t) = \infty$ and $\sum_{t=0}^{\infty}\alpha(t)^2 < \infty$, we have $\lim_{T\to\infty}\mathbb{E}(f(\hat{\boldsymbol{x}}(T)) - f(\boldsymbol{x}^*)) = 0$, which implies that $\hat{\boldsymbol{x}}(T)$ approaches the optimal solution.

B. Resource Allocation

Based on Theorem 1, problem (19) can be reformulated as

$$\min_{\boldsymbol{A},\boldsymbol{p}} \quad \frac{R^2 + \sum_{n=1}^{N} d_1(1-q_n)}{d_2(1-\max_{n\in\mathcal{N}} q_n)} \tag{22}$$

$$\text{s.t.} \quad (19a) - (19e), \tag{22a}$$

To solve problem (22), we first analyze the optimal condition.

Lemma 2. For the optimal solution of problem (22), we have

$$q_1^* = \cdots = q_N^*. \tag{23}$$

Proof: Assume that the optimal solution of problem (22) is $(\boldsymbol{A}^*, \boldsymbol{p}^*)$ and there exist i and j such that $q_i^* < q_j^*$. We construct a new solution $(\boldsymbol{A}^*, \bar{\boldsymbol{p}})$ with

$$\bar{p}_n = p_n^*,\ \bar{p}_i = p_i^* - \epsilon,\ \forall n \neq i, \tag{24}$$

where $\epsilon > 0$ is a small positive constant that satisfies $q_i^* < \bar{q}_i < q_j^*$. Since the new solution $(\boldsymbol{A}^*, \bar{\boldsymbol{p}})$ is feasible and has lower objective value compared to the solution $(\boldsymbol{A}^*, \boldsymbol{p}^*)$, which contradicts that solution $(\boldsymbol{A}^*, \boldsymbol{p}^*)$ is optimal. As a result, the optimal condition (23) always holds for problem (22). ■

Based on Lemma 2, the objective function in (22) is equivalent to $\frac{R^2+\sum_{n=1}^{N} d_1(1-q_n)}{d_2(1-\max_{n\in\mathcal{N}} q_n)} = \frac{R^2}{d_2(1-\max_{n\in\mathcal{N}} q_n)} + Nd_1$. Meanwhile, minimizing $\frac{R^2}{d_2(1-\max_{n\in\mathcal{N}} q_n)}$ is equal to minimize $\max_{n\in\mathcal{N}} q_n$. Introducing a new variable $q = \max_{n\in\mathcal{N}} q_n$, problem (22) can be simplified as

$$\min_{\boldsymbol{A},\boldsymbol{p},q} \quad q \tag{25}$$

$$\text{s.t.} \quad q \geq 1 - \sum_{l=1}^{n} a_{ln}\exp\left(-\frac{D_{ln}}{p_n}\right), \quad \forall n \in \mathcal{N}, \tag{25a}$$

$$(19a) - (19e), \tag{25b}$$

where inequality (25a) holds with equality for the optimal solution as otherwise the objective value can be further improved. To solve problem (25), we use an iterative method, which optimizes $\boldsymbol{A}$ and $\boldsymbol{p}$ in an alternating manner.

Give power vector $\boldsymbol{p}$, problem (25) is a mixed linear integer problem. By temporally relaxing integer variable $a_{ln} \in [0, 1]$, problem (25) with fixed $\boldsymbol{p}$ is a standard linear problem, which can be effectively solved via the simplex method. Then, we can obtain the integer value of a_{ln} by using the rounding method.

Algorithm 2 Iterative RB Allocation and Power Control

1: Initialize RB allocation $\boldsymbol{A}$ and power contrl $\boldsymbol{p}$.
2: **repeat**
3: With fixed power control $\boldsymbol{p}$, optimize RB allocation with the simplex method and rounding technique.
4: With fixed RB association $\boldsymbol{A}$, obtain the optimal $\boldsymbol{p}$ by solving (27) and (28).
5: **until** the objective value (25) converges.

With fixed RB association $\boldsymbol{A}$, problem (25) reduces to

$$\min_{\boldsymbol{p},q} \quad q \tag{26}$$

$$\text{s.t.} \quad q \geq 1 - \exp\left(-\frac{D_{l_n n}}{p_n}\right), \quad \forall n \in \mathcal{N}, \tag{26a}$$

$$\sum_{n=1}^{N} p_n \leq P_{\max}, \tag{26b}$$

$$0 \leq p_n \leq P_n, \quad \forall n \in \mathcal{N}, \tag{26c}$$

where l_n is the assigned RB for user n, i.e., $a_{l_n n} = 1$. According to Lemma 2, constraint (26a) holds with equality for the optimal solution and we can obtain

$$p_n^* = -\frac{D_{l_n n}}{\ln(1-q^*)}. \tag{27}$$

Substituting p_n^* into constraints (26a)-(26b), the optimal q^* should satisfy

$$\sum_{n=1}^{N} -\frac{D_{l_n n}}{\ln(1-q^*)}\Bigg|^{P_n} \leq P_{\max}, \tag{28}$$

where $a|^b = \min\{a, b\}$. Since the left hand-side of (28) is a decreasing function with respect to q^*, the minimal q^* satisfying (28) can be obtained by using the bisection method.

C. Complexity Analysis

The iterative algorithm for solving (22) is given in Algorithm 2. The major complexity in each iteration lies in solving the RB allocation subproblem and the power control subproblem. With fixed power control, the complexity of using the simplex method is $\mathcal{O}(N^3)$ [14] for solving (25). With fixed RB association, the complexity of solving (28) with the bisection method is $\mathcal{O}(N\log(1/\epsilon))$, where ϵ is the accuracy of the bisection method. As a result, the total complexity of Algorithm 2 is $\mathcal{O}(T_0N^3 + T_0N\log(1/\epsilon))$, where T_0 is the number of iterations in Algorithm 2.

IV. Simulation Results

There are $N = 50$ users uniformly in a square area of size 500 m × 500 m with the BS at the center. The path loss model is $128.1+37.6\log_{10} d$ (d is in km). The bandwidth of each RB is 1 MHz and the noise power spectral density is $N_0 = -174$ dBm/Hz. The maximum transmit power of each user is set as $P_n = 10$ dBm. To show the performance of the primal-dual algorithm, we consider the similar parameters as in [10].

The convergence of the distributed primal-dual algorithm is shown in Fig. 1. In the figure, we compare the proposed algorithm with the conventional algorithm which ignores the

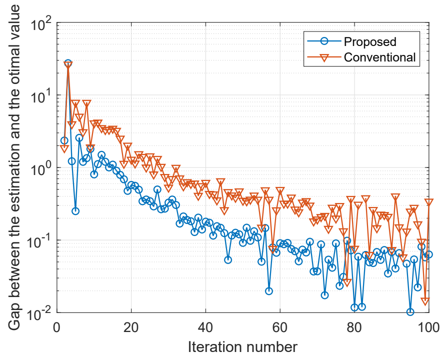

Fig. 1. Convergence behaviour of the distributed primal-dual algorithm.

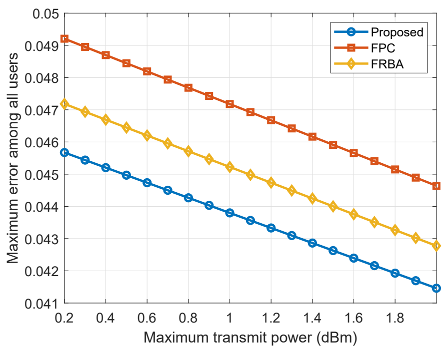

Fig. 2. Maximum transmission error among all user versus the maximum sum transmit power.

wireless affect. For the conventional algorithm, each user transmits with equal transmit power and RB allocation is randomly assigned. From this figure, we find that the distributed primal-dual algorithm has an oscillatory behavior. It is found that the proposed algorithm achieves up to 52% gap reduction compared to the conventional algorithm.

We compare the proposed Algorithm 2 to solve problem (25) with two baselines: the fixed power control algorithm with only optimizing RB allocation (labelled as 'FPC') and the fixed RB allocation algorithm with only optimizing power control (labelled as 'FRBA'). Fig. 2 illustrates the maximum transmission error among all user versus the maximum sum transmit power. From this figure, the maximum transmission error decreases for all schemes as the maximum sum transmit power varies. This is because high transmit power can decrease the transmission error. It is observed that the proposed algorithm achieves the best performance, which shows the superiority of the joint RB allocation and power control design.

V. Conclusions

In this paper, we have investigated the convergence optimization problem of a distributed primal-dual algorithm over wireless communication networks via jointly optimizing RB allocation and power control. We have derived a closed-form expression for the expected convergence rate of a distributed primal-dual algorithm that considers the transmission errors over wireless communications. Based on this convergence rate, we have first obtained an optimal condition for the resource allocation. Then, an iterative algorithm has been proposed, where a closed-form solution has been obtained for the power control subproblem. Simulation results have shown the superiority of the proposed solution.

Appendix A
Proof of Theorem 1

To prove Theorem 1, we first obtain the minimum mean square error between $\boldsymbol{x}(t+1)$ and $\boldsymbol{x}^*$, i.e.,

$$\begin{aligned}
&\mathbb{E}\|\boldsymbol{x}(t+1)-\boldsymbol{x}^*\|^2\\
&\overset{(17)}{=}\mathbb{E}\left\|\frac{\sum_{n=1}^N C_n(t)\boldsymbol{y}_n(t+1)}{\sum_{n=1}^N C_n(t)}-\boldsymbol{x}^*\right\|^2\\
&\overset{(8)}{=}\mathbb{E}\left\|\frac{\sum_{n=1}^N C_n(t)\left(\boldsymbol{x}(t)-\alpha(t)\nabla_{\boldsymbol{x}}\mathcal{L}_n(\boldsymbol{x}(t),\boldsymbol{\lambda}(t))\right)}{\sum_{n=1}^N C_n(t)}-\boldsymbol{x}^*\right\|^2\\
&\overset{(a)}{\leq}\sum_{n=1}^N\mathbb{E}\left(\frac{C_n(t)}{\sum_{i=1}^N C_i(t)}\right)\|\boldsymbol{x}(t)-\alpha(t)\nabla_{\boldsymbol{x}}\mathcal{L}_n(\boldsymbol{x}(t),\boldsymbol{\lambda}(t))-\boldsymbol{x}^*\|^2\\
&=\|\boldsymbol{x}(t)-\boldsymbol{x}^*\|^2+\sum_{n=1}^N\mathbb{E}\left(\frac{C_n(t)}{\sum_{i=1}^N C_i(t)}\right)\\
&\quad\times\Big(\alpha(t)^2\|\nabla_{\boldsymbol{x}}\mathcal{L}_n(\boldsymbol{x}(t),\boldsymbol{\lambda}(t))\|^2\\
&\quad-2\alpha(t)(\boldsymbol{x}(t)-\boldsymbol{x}^*)^T\nabla_{\boldsymbol{x}}\mathcal{L}_n(\boldsymbol{x}(t),\boldsymbol{\lambda}(t))\Big),
\end{aligned}\tag{A.1}$$

where inequality (a) follows from the fact that squared norm is a convex function. To obtain an upper bound of $\mathbb{E}\left(\frac{C_n(t)}{\sum_{i=1}^N C_i(t)}\right)$, we define $\kappa_n=\frac{C_n(t)}{\sum_{i=1}^N C_i(t)}$. Based on (18), we have

$$\kappa_n=\begin{cases}\frac{1}{1+\sum_{i=1,i\neq n}^N C_i(t)}, & \text{with probability } 1-q_n\\ 0, & \text{with probability } q_n\end{cases}\tag{A.2}$$

Since $\frac{1}{N}\leq\frac{1}{1+\sum_{i=1,i\neq n}^N C_i(t)}\leq 1$, we can obtain $\mathbb{E}(\kappa_n)\leq 1-q_n$. Combining (A.1) and $\mathbb{E}(\kappa_n)\leq 1-q_n$ yields

$$\begin{aligned}
&\mathbb{E}\|\boldsymbol{x}(t+1)-\boldsymbol{x}^*\|^2\\
&\leq\|\boldsymbol{x}(t)-\boldsymbol{x}^*\|^2+\sum_{n=1}^N(1-q_n)\Big(\alpha(t)^2\|\nabla_{\boldsymbol{x}}\mathcal{L}_n(\boldsymbol{x}(t),\boldsymbol{\lambda}(t))\|^2\\
&\quad-2\alpha(t)(\boldsymbol{x}(t)-\boldsymbol{x}^*)^T\nabla_{\boldsymbol{x}}\mathcal{L}_n(\boldsymbol{x}(t),\boldsymbol{\lambda}(t))\Big).
\end{aligned}\tag{A.3}$$

According to the recursion in (A.3) with $\boldsymbol{x}(0)=\mathbf{0}$, we have

$$\begin{aligned}
&\mathbb{E}\|\boldsymbol{x}(T)-\boldsymbol{x}^*\|^2\\
&\leq\|\boldsymbol{x}^*\|^2+\sum_{t=0}^{T-1}\sum_{n=1}^N(1-q_n)\Big(\alpha(t)^2\|\nabla_{\boldsymbol{x}}\mathcal{L}_n(\boldsymbol{x}(t),\boldsymbol{\lambda}(t))\|^2\\
&\quad-2\alpha(t)(\boldsymbol{x}(t)-\boldsymbol{x}^*)^T\nabla_{\boldsymbol{x}}\mathcal{L}_n(\boldsymbol{x}(t),\boldsymbol{\lambda}(t))\Big).
\end{aligned}\tag{A.4}$$

Due to the non-negativity of left-hand side in (A.4), we have

$$\begin{aligned}
&\|\boldsymbol{x}^*\|^2+\sum_{t=0}^{T-1}\sum_{n=1}^N(1-q_n)\alpha(t)^2\|\nabla_{\boldsymbol{x}}\mathcal{L}_n(\boldsymbol{x}(t),\boldsymbol{\lambda}(t))\|^2\geq\\
&2\sum_{t=0}^{T-1}\sum_{n=1}^N(1-q_n)\alpha(t)(\boldsymbol{x}(t)-\boldsymbol{x}^*)^T\nabla_{\boldsymbol{x}}\mathcal{L}_n(\boldsymbol{x}(t),\boldsymbol{\lambda}(t)).
\end{aligned}\tag{A.5}$$

According to Assumption 3, $\mathcal{L}_n(\boldsymbol{x},\boldsymbol{\lambda})$ is convex with respect to $\boldsymbol{x}$. Hence, we have

$$(\boldsymbol{x}(t)-\boldsymbol{x}^*)^T\nabla_{\boldsymbol{x}}\mathcal{L}_n(\boldsymbol{x}(t),\boldsymbol{\lambda}(t)) \geq \mathcal{L}_n(\boldsymbol{x}(t),\boldsymbol{\lambda}(t))-\mathcal{L}_n(\boldsymbol{x}^*,\boldsymbol{\lambda}(t)). \tag{A.6}$$

Based on (A.5) and (A.6), we have

$$\begin{aligned}
&\|\boldsymbol{x}^*\|^2+\sum_{t=0}^{T-1}\sum_{n=1}^{N}(1-q_n)\alpha(t)^2\|\nabla_{\boldsymbol{x}}\mathcal{L}_n(\boldsymbol{x}(t),\boldsymbol{\lambda}(t)\|^2\\
\geq&2\sum_{t=0}^{T-1}\sum_{n=1}^{N}(1-q_n)\alpha(t)(\mathcal{L}_n(\boldsymbol{x}(t),\boldsymbol{\lambda}(t))-\mathcal{L}_n(\boldsymbol{x}^*,\boldsymbol{\lambda}(t)))\\
\overset{(3)}{=}&2\sum_{t=0}^{T-1}\sum_{n=1}^{N}(1-q_n)\alpha(t)(f_n(\boldsymbol{x}(t))-f_n(\boldsymbol{x}^*))\\
&+2\sum_{t=0}^{T-1}\sum_{n=1}^{N}(1-q_n)\alpha(t)(\boldsymbol{\lambda}(t)^T\boldsymbol{g}(\boldsymbol{x}(t))-\boldsymbol{\lambda}(t)^T\boldsymbol{g}(\boldsymbol{x}^*))\\
\overset{(b)}{\geq}&2\sum_{t=0}^{T-1}\sum_{n=1}^{N}(1-q_n)\alpha(t)(f_n(\boldsymbol{x}(t))-f_n(\boldsymbol{x}^*))\\
&+2\sum_{t=0}^{T-1}\sum_{n=1}^{N}(1-q_n)\alpha(t)\boldsymbol{\lambda}(t)^T\boldsymbol{g}(\boldsymbol{x}(t)),
\end{aligned} \tag{A.7}$$

where (b) follows from the fact that $g_m(\boldsymbol{x}^*)\leq 0$ and $\lambda_m(t)\geq 0$. To derive a lower bound for the last term in the right hand side of (A.7), we provide the following lemma.

Lemma 3. For all T, the following inequality holds

$$\begin{aligned}
&2\sum_{t=0}^{T-1}\sum_{n=1}^{N}(1-q_n)\alpha(t)\boldsymbol{\lambda}(t)^T\boldsymbol{g}(\boldsymbol{x}(t))\\
&\geq-\sum_{t=0}^{T-1}\sum_{n=1}^{N}(1-q_n)\alpha(t)^2\|\boldsymbol{g}(\boldsymbol{x}(t))\|^2.
\end{aligned} \tag{A.8}$$

Proof: According to (7), we have

$$\begin{aligned}
&\|\boldsymbol{\lambda}(t+1)\|^2\overset{(7)}{=}\|\boldsymbol{\lambda}(t)+\alpha(t)\boldsymbol{g}(\boldsymbol{x}(t))\|^2\\
=&\|\boldsymbol{\lambda}(t)\|^2+2\alpha(t)(\boldsymbol{\lambda}(t))^T\boldsymbol{g}(\boldsymbol{x}(t))+\alpha(t)^2\|\boldsymbol{g}(\boldsymbol{x}(t))\|^2.
\end{aligned}$$

By using the recursion method and $\boldsymbol{\lambda}(0)=\boldsymbol{0}$, we can obtain

$$\|\boldsymbol{\lambda}(T)\|^2=2\sum_{t=0}^{T-1}\alpha(t)(\boldsymbol{\lambda}(t))^T\boldsymbol{g}(\boldsymbol{x}(t))+\sum_{t=0}^{T-1}\alpha(t)^2\|\boldsymbol{g}(\boldsymbol{x}(t))\|^2.$$

Since $\|\boldsymbol{\lambda}(T)\|^2\geq 0$, we can obtain (A.8). ■

Using Lemma 3, (A.8) can be rewritten as

$$\begin{aligned}
&\|\boldsymbol{x}^*\|^2+\sum_{t=0}^{T-1}\sum_{n=1}^{N}(1-q_n)\alpha(t)^2\|\nabla_{\boldsymbol{x}}\mathcal{L}_n(\boldsymbol{x}(t),\boldsymbol{\lambda}(t)\|^2\\
\overset{(c)}{\geq}&2N\sum_{t=0}^{T-1}(1-q_0)\alpha(t)(f(\boldsymbol{x}(t))-f(\boldsymbol{x}^*))\\
&-\sum_{t=0}^{T-1}\sum_{n=1}^{N}(1-q_n)\alpha(t)^2\|\boldsymbol{g}(\boldsymbol{x}(t))\|^2,
\end{aligned} \tag{A.9}$$

where (c) follows from the definition $q_0=\max_{n\in\mathcal{N}}q_n$ and $f(\boldsymbol{x})=\frac{1}{N}\sum_{n=1}^{N}f_n(\boldsymbol{x})$.

Recall the definition of $\hat{\boldsymbol{x}}(T)$ in (10), we have

$$\begin{aligned}
&\mathbb{E}(f(\hat{\boldsymbol{x}}(T))-f(\boldsymbol{x}^*))\overset{(d)}{\leq}\frac{\sum_{t=0}^{T-1}\alpha(t)(f(\boldsymbol{x}(t))-f(\boldsymbol{x}^*))}{\sum_{t=0}^{T-1}\alpha(t)}\\
\overset{(A.9)}{\leq}&\frac{1}{2N(1-q_0)\sum_{t=0}^{T-1}\alpha(t)}\Big(\|\boldsymbol{x}^*\|^2\\
&+\sum_{t=0}^{T-1}\sum_{n=1}^{N}(1-q_n)\alpha(t)^2\|\nabla_{\boldsymbol{x}}\mathcal{L}_n(\boldsymbol{x}(t),\boldsymbol{\lambda}(t)\|^2\\
&+\sum_{t=0}^{T-1}\sum_{n=1}^{N}(1-q_n)\alpha(t)^2\|\boldsymbol{g}(\boldsymbol{x}(t))\|^2\Big)\\
\overset{(e)}{\leq}&\frac{1}{2N(1-q_0)\sum_{t=0}^{T-1}\alpha(t)}\Big(\|\boldsymbol{x}^*\|^2\\
&+\sum_{t=0}^{T-1}\sum_{n=1}^{N}(1-q_n)\alpha(t)^2\|\nabla_{\boldsymbol{x}}\mathcal{L}_n(\boldsymbol{x}(t),\boldsymbol{\lambda}(t)\|^2\\
&+\sum_{t=0}^{T-1}\sum_{n=1}^{N}(1-q_n)\alpha(t)^2ML^2R^2\Big),
\end{aligned} \tag{A.10}$$

where (d) follows from the convexity of function $f(\boldsymbol{x})$ and (e) follows from Assumption 3. To derive an upper bound for $\|\nabla_{\boldsymbol{x}}\mathcal{L}_n(\boldsymbol{x}(t),\boldsymbol{\lambda}(t)\|$, we have

$$\begin{aligned}
&\|\nabla_{\boldsymbol{x}}\mathcal{L}_n(\boldsymbol{x}(t),\boldsymbol{\lambda}(t)\|\overset{(3)}{=}\|\nabla f_n(\boldsymbol{x}(t))+\boldsymbol{\lambda}(t)^T\nabla\boldsymbol{g}(\boldsymbol{x}(t))\|\\
&\overset{(f)}{\leq}L(1+\|\boldsymbol{\lambda}(t)\|_1)\overset{(g)}{\leq}L(1+MS),
\end{aligned} \tag{A.11}$$

where (f) and (g) Assumptions 3 and 2, respectively.

Based on (A.11), (21) can be derived from (A.10).

References

[1] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," *IEEE network*, vol. 34, no. 3, pp. 134–142, 2019.

[2] M. Chen, Z. Yang, W. Saad, C. Yin, H. V. Poor, and S. Cui, "A joint learning and communications framework for federated learning over wireless networks," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 269–283, Jan. 2021.

[3] Z. Yang, M. Chen, W. Saad, C. S. Hong, and M. Shikh-Bahaei, "Energy efficient federated learning over wireless communication networks," *arXiv preprint arXiv:1911.02417*, 2019.

[4] Z. Yang, M. Chen, K.-K. Wong, H. V. Poor, and S. Cui, "Federated learning for 6g: Applications, challenges, and opportunities," *arXiv preprint arXiv:2101.01338*, 2021.

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

[6] A. Nedić and A. Olshevsky, "Distributed optimization over time-varying directed graphs," *IEEE Trans. Autom. Control*, vol. 60, no. 3, pp. 601–615, 2014.

[7] W. Shi, Q. Ling, G. Wu, and W. Yin, "A proximal gradient algorithm for decentralized composite optimization," *IEEE Trans. Signal Process.*, vol. 63, no. 22, pp. 6013–6023, 2015.

[8] K. Yuan, Q. Ling, and W. Yin, "On the convergence of decentralized gradient descent," *SIAM J. Optim.*, vol. 26, no. 3, pp. 1835–1854, 2016.

[9] J. F. Mota, J. M. Xavier, P. M. Aguiar, and M. Püschel, "D-ADMM: A communication-efficient distributed algorithm for separable optimization," *IEEE Trans. Signal Process.*, vol. 61, no. 10, pp. 2718–2723, 2013.

[10] M. B. Khuzani and N. Li, "Distributed regularized primal-dual method: Convergence analysis and trade-offs," *arXiv preprint arXiv:1609.08262*, 2016.

[11] Y. Liu, W. Xu, G. Wu, Z. Tian, and Q. Ling, "Communication-censored ADMM for decentralized consensus optimization," *IEEE Trans. Signal Process.*, vol. 67, no. 10, pp. 2565–2579, 2019.

[12] A. Elgabli, J. Park, A. S. Bedi, M. Bennis, and V. Aggarwal, "Communication efficient framework for decentralized machine learning," in *Proc. Annual Conference on Information Sciences and Systems (CISS)*, 2020, pp. 1–5.

[13] H. Chen, Y. Ye, M. Xiao, M. Skoglund, and H. V. Poor, "Coded stochastic ADMM for decentralized consensus optimization with edge computing," *arXiv preprint arXiv:2010.00914*, 2020.

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