Device Selection of Distributed Primal-Dual Algorithms Over Wireless Networks

Zhaohui Yang*, Chongwen Huang†, Hao Xu‡, Wei Xu§, Yue Cao¶
*Department of Engineering, University College London, WC1E 6BT, UK.
†College of Information Science and Electronic Engineering Zhejiang Key Lab
of Information Processing Communication and Networking, Zhejiang University, China.
‡Faculty of Electrical Engineering and Computer Science, Technical University of Berlin, Germany.
§National Mobile Communications Research Laboratory, Southeast University, China.
¶School of Cyber Science and Engineering, Wuhan University, China.
Emails: yang.zhaohui@kcl.ac.uk, chongwenhuang@zju.edu.cn, xuhao@mail.tu-berlin.de, wxu@seu.edu.cn, yue.cao@whu.edu.cn

2021 IEEE 94th Vehicular Technology Conference (VTC2021-Fall) | 978-1-6654-1368-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/VTC2021-FALL52928.2021.9625364

***Abstract*—In this paper, the implementation of a distributed primal-dual learning algorithm over realistic wireless networks is investigated. In the considered model, the users and one base station (BS) cooperatively perform a distributed primal-dual learning 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 provided in a closed form while considering the impact of wireless factors such as data transmission errors. Simulation results show that the proposed distributed primal-dual algorithm can reduce the gap between the target and obtained solution compared to the distributed primal-dual learning algorithm without considering imperfect wireless transmission.**

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

I. Introduction

Due to the explosive growth in data traffic, machine learning and data driven approaches have recently received much attention and are anticipated to be a key enabler for the to be developed sixth generation (6G) wireless networks [1] including vehicular to everything networks. Nowadays, standard machine learning approaches require centralizing the training data on a single data center or cloud. Since massive data samples need to be uploaded to the data center, transmission delay can be very high and user privacy is not guaranteed in standard centralized machine learning approaches. However, low-latency and privacy requirements are important in the emerging application scenarios, such as unmanned aerial vehicles, extended reality (XR) services, autonomous driving, which makes centralized machine learning approaches inapplicable. Moreover, due to limited communication resources, it is impractical for all the wireless devices that are engaged in learning to transmit all of their collected data to a data center that uses a centralized learning algorithm for data analytic or network self-organization [1]–[11].

Therefore, it becomes increasingly attractive to process data locally at edge devices. This has led to the emergency of distributed optimization methods. In distributed optimization, each node can compute on its own data and sends the results to its neighbours or a central node. Distributed optimization has many applications, such as user selection optimization, resource allocation optimization, trajectory optimization, and distributed machine learning design [12].

Distributed optimization algorithms fall within two main classes: distributed primal algorithms [13], [14] and distributed primal-dual algorithms [15]. Combining the advantages of distributed optimization and machine learning, distributed learning frameworks are needed to enable wireless devices to collaboratively build a shared learning model with training taken place locally. One of the most promising distributed learning algorithms is the emerging federated learning [16], [17] framework is anticipated in future Internet of Things (IoT) systems. In federated learning, wireless devices can cooperatively execute a learning task by only uploading local learning models to the base station (BS) instead of sharing the entirety of their training data. Since the data center cannot access the local data sets at the users, distributed machine learning can protect data privacy of the users. Using gradient sparsification, a digital transmission scheme based on gradient quantization was investigated in [16]. To implement federated learning over wireless networks, the wireless devices must transmit their local training results over wireless links [17], which can affect the performance of federated learning due to limited wireless resources (such as time and bandwidth). Due to the advantage of fast convergence for primal distributed primal-dual algorithms, it is meaningful to investigate the combination of distributed learning and primal-dual algorithms. However,

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.

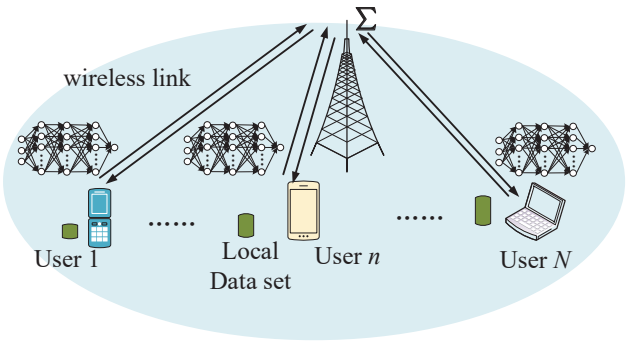

Fig. 1. A distributed learning algorithm over wireless communication systems.

the above works [16], [17] ignored considering distributed learning and primal-dual algorithms with taking transmission error into consideration.

The main contribution of this paper is a novel framework that enables the implementation of a primal-dual learning algorithm over a realistic wireless network. Our key contributions include:

- We study the performance of the distributed primal-dual learning algorithm over wireless communication networks. For the considered primal-dual learning algorithm, we provide the convergence rate while considering the impact of wireless factors such as data transmission error.
- 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 learning problem with the primal-dual algorithm, as shown in Fig. 1. 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 machine learning problem [18], [19]:

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

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

where $f_n(\boldsymbol{x}, \mathcal{D}_n)$ is the loss function, $g_m(\boldsymbol{x})$ is the constraint function, $\mathcal{M} = \{1, \cdots, M\}$, and M is the number of constraints[1]. The optimization variable $\boldsymbol{x}$ stands for the weight vector of the machine learning problem. 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

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

Based on the definition of the local Lagrange function, the distributed primal-dual learning algorithm is proposed to solve the following maximin problem [19]:

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

The distributed primal-dual learning 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 The total number of RBs is D. Let $b_{ln} \in \{0, 1\}$ denote the RB association index, i.e., $b_{ln} = 1$ implies that RB l is assigned to user n and $b_{ln} = 0$ otherwise. Since each user can occupy at most one RB and each RB should be occupied by only one user, we have

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

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

[1]For different learning tasks, the loss function will be different. For example, quadric function for linear regression and log function for logistic regression. For the constraints, $g_m(\boldsymbol{x})$ can be box constraints in the logistic regression.

Algorithm 1 Distributed Primal-Dual Learning Algorithm

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

$$\boldsymbol{\lambda}(t+1) = \boldsymbol{\lambda}(t) + \alpha(t)\nabla_{\boldsymbol{\lambda}}\mathcal{L}_n(\boldsymbol{x}(t), \boldsymbol{\lambda}(t)), \quad (7)$$

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

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

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

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

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

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

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

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

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

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

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

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

$$q_n = \sum_{l=1}^{D} b_{ln} q_{ln}. \quad (16)$$

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[2]

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

where

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

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

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

$$\text{s.t.} \quad \sum_{l=1}^{D} b_{ln} \le 1, \quad \forall l \in \mathcal{N}, \quad (19a)$$

$$\sum_{n=1}^{N} b_{ln} = 1, \quad \forall n \in \mathcal{N}, \quad (19b)$$

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

$$b_{ln} \in \{0, 1\}, \quad \forall l, n \in \mathcal{N}, \quad (19d)$$

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

where $\boldsymbol{B} = \{b_{ln}\}_{N\times N}$, $\boldsymbol{p} = [p_1, \cdots, p_N]^T$, $\mathbb{E}(f(\hat{\boldsymbol{x}}(T)) - f(\boldsymbol{x}^*))$ 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. Algorithm Design

A. RB Allocation

In practical scenarios, such as IoT systems, there are a large number of users. Due to the limited bandwidth resource for wireless communications, we assume that the total number of

[2]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.

Algorithm 2 Dynamic User Association Scheme

1: **Input**: RB allocation probability matrix $\boldsymbol{B} = \{b_{ln}\}_{D\times N}$.
2: Initialize the set of users with assigned RB $\mathcal{E} = \emptyset$ and the set of users without any associated RB $\mathcal{F} = \mathcal{N}$.
3: **for** $l = 1:1:D$ **do**
4: Uniformly generate a variable u in interval $[0, 1]$.
5: Scale the probability of users without any associated RB as

$$b_{ln} = \frac{b_{ln}}{\sum_{s\in\mathcal{F}} b_{sn}}, \quad \forall l \in \mathcal{F}. \tag{20}$$

6: Calculate the user id n_l for occupying RB l, which should satisfy

$$\sum_{s<n_l, s\in\mathcal{F}} b_{sn} \le u \le \sum_{s\le n_l, s\in\mathcal{F}} b_{sn}. \tag{21}$$

7: Update $\mathcal{E} = \mathcal{E} \cup \{n_l\}$ and $\mathcal{F} = \mathcal{F} \setminus \{n_l\}$.
8: **end for**

RBs is D, which is smaller than the total number of users, i.e., $D < N$. In this case, all users cannot be served at the same time. To ensure that all users can be involved the distributed primal-dual learning framework, we consider the dynamic user association scheme. The basic idea is that only a small number of users can be served under a given probability model at each time and all users can be involved during the whole distributed primal-dual learning process. To be specific, denote b_{ln} as the probability that RB l is allocated to user n. Due to that fact that each RB can be occupied by only one user during each transmission, we have the following constraint

$$\sum_{n=1}^{N} b_{ln} = 1, \quad \forall l \in \mathcal{D}, \tag{22}$$

where $\mathcal{D} = \{1, 2, \cdots, D\}$ is the set of all RBs.

With given RB allocation probability matrix $\boldsymbol{B} = \{b_{ln}\}_{D\times N}$, the dynamic user association scheme is presented in Algorithm 2. In Algorithm 2, each RB is assigned iteratively as shown at step 3. At step 4 in Algorithm 2, a random variable is generated to determine one user, which can be assigned with RB l. The user is selected based on the RB allocation probability matrix $\boldsymbol{B}$ according to step 5. The sets of users with and without assigned RB are updated at step 6.

B. Convergence Analysis

Problem (19) is hard to solve since the accurate formulation of the objective function is difficult to derive. In this section, we provide the convergence analysis of the proposed distributed primal-dual learning Algorithm 1, which is helpful in simplifying the objective function in problem (19).

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 $g_m(\boldsymbol{x}) < 0, \forall m \in \mathcal{M}$. Further assume that $g_m(\boldsymbol{0}) = 0, \forall m \in \mathcal{M}$.

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.,

$$\|\nabla f_n(\boldsymbol{x})\| \le L, \|\nabla g_m(\boldsymbol{x})\| \le L, \quad \forall n \in \mathcal{N}, m \in \mathcal{M}, \tag{23}$$

where $L < \infty$ is a constant.

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

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

$$\mathbb{E}(f(\hat{\boldsymbol{x}}(T)) - f(\boldsymbol{x}^*)) \le \frac{R^2 + \sum_{n=1}^{N} d_1(1-q_n)}{d_2(1-q_0)} \tag{24}$$

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

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.

Theorem 2. If the BS and the users implement Algorithm 1 over T iterations and the step size $\alpha(t)$ is chosen as $\alpha(t) = \frac{R}{\sqrt{1+t}}$, the upper bound of $\mathbb{E}(f(\hat{\boldsymbol{x}}(T)) - f(\boldsymbol{x}^*))$ can be given by:

$$\begin{aligned}&\mathbb{E}(f(\hat{\boldsymbol{x}}(T)) - f(\boldsymbol{x}^*))\\ &\le \frac{R(1+\sum_{n=1}^{N}(1-q_n)L^2((1+MS)^2+MR^2)\ln T)}{2(1-q_0)\sqrt{T}}.\end{aligned} \tag{25}$$

According to the upper bound (25), the convergence rate of the Algorithm 1 is given by $\mathcal{O}(T^{-\frac{1}{2}})$. Theorem 2 implies that the obtained solution approaches the optimal solution as the number of iterations grows.

IV. Simulation Results

There are $N = 100$ users uniformly in a square area of size 500 m $\times$ 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 $P_n = 5$ dBm. To show the performance of the distributed primal-dual learning algorithm, we consider

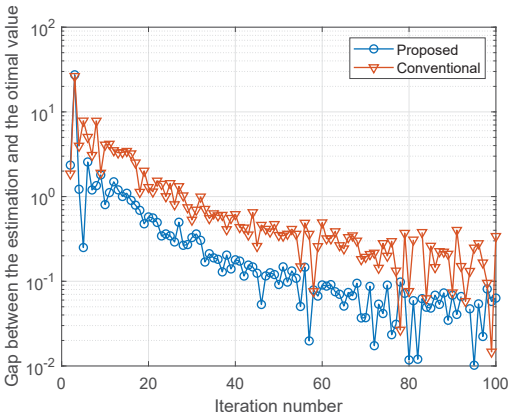

Fig. 2. Convergence behaviour of the distributed primal-dual learning algorithm.

the similar parameters as in [19]. In particular, each user trains the learning model using the MNIST dataset.

The convergence of the distributed primal-dual learning algorithm is shown in Fig. 2. In the figure, we compare the proposed algorithm with the conventional algorithm which ignores the 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 learning algorithm has an oscillatory behavior.

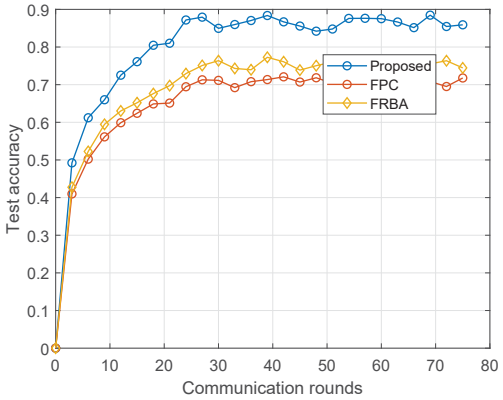

Fig. 3. Test accuracy versus the the number of communication rounds with $N = 100$ and $P_n = 1$ dBm.

We compare the proposed algorithm 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').

Figs. 3 shows the test accuracy versus the number of communication rounds. From this figure, it is found that the test accuracy has an increasing trend. It is also found that the proposed algorithm achieves the highest test probability.

The proof of Theorems 1 and 2 can be similarly derived as in [19]. Based on Theorem 2, the objective function of problem (19) can be approximated. Then, an iterative algorithm can be used to solve problem (19) by optimizing $\boldsymbol{A}$ and $\boldsymbol{p}$ iteratively.

V. Conclusions

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

References

[1] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," *IEEE Network*, 2020 (To appear).

[2] J. Park, S. Samarakoon, M. Bennis, and M. Debbah, "Wireless network intelligence at the edge," *Proceedings of the IEEE*, vol. 107, no. 11, pp. 2204–2239, Nov. 2019.

[3] M. Chen, O. Semiari, W. Saad, X. Liu, and C. Yin, "Federated echo state learning for minimizing breaks in presence in wireless virtual reality networks," *IEEE Trans. Wireless Commun.*, vol. 19, no. 1, pp. 177–191, Jan. 2020.

[4] H. Tembine, *Distributed strategic learning for wireless engineers*. CRC Press, 2018.

[5] S. Samarakoon, M. Bennis, W. Saad, and M. Debbah, "Distributed federated learning for ultra-reliable low-latency vehicular communications," *arXiv preprint arXiv:1807.08127*, 2018.

[6] D. Gündüz, P. de Kerret, N. D. Sidiropoulos, D. Gesbert, C. R. Murthy, and M. van der Schaar, "Machine learning in the air," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 10, pp. 2184–2199, Oct. 2019.

[7] 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.

[8] M. S. H. Abad, E. Ozfatura, D. Gündüz, and O. Ercetin, "Hierarchical federated learning across heterogeneous cellular networks," *arXiv preprint arXiv:1909.02362*, 2019.

[9] 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.

[10] T. Nishio and R. Yonetani, "Client selection for federated learning with heterogeneous resources in mobile edge," in *Proc. IEEE Int. Conf. Commun.* Shanghai, China: IEEE, May 2019, pp. 1–7.

[11] S. P. Karimireddy, S. Kale, M. Mohri, S. J. Reddi, S. U. Stich, and A. T. Suresh, "SCAFFOLD: Stochastic controlled averaging for on-device federated learning," *arXiv preprint arXiv:1910.06378*, 2019.

[12] 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," *arXiv preprint arXiv:1909.07972*, 2019.

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

[14] 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.

[15] 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.

[16] M. M. Amiri and D. Gündüz, "Machine learning at the wireless edge: Distributed stochastic gradient descent over-the-air," in *Proc. IEEE Int. Symp. Information Theory*, Paris, France, Jan. 2019, pp. 1432–1436.

[17] G. Zhu, D. Liu, Y. Du, C. You, J. Zhang, and K. Huang, "Towards an intelligent edge: Wireless communication meets machine learning," *arXiv preprint arXiv:1809.00343*, 2018.

[18] H. H. Yang, Z. Liu, T. Q. S. Quek, and H. V. Poor, "Scheduling policies for federated learning in wireless networks," *IEEE Trans. Commun.*, vol. 68, no. 1, pp. 317–333, Jan. 2020.

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