

Decentralized Consensus Optimization Based on Parallel Random Walk

Yu Ye[✉], *Student Member, IEEE*, Hao Chen, Zheng Ma, *Member, IEEE*, and Ming Xiao[✉], *Senior Member, IEEE*

Abstract—The alternating direction method of multipliers (ADMM) has recently been recognized as a promising approach for large-scale machine learning models. However, very few results study ADMM from the aspect of communication costs, especially jointly with running time. In this letter, we investigate the communication efficiency and running time of ADMM in solving the consensus optimization problem over decentralized networks. We first review the effort of random walk ADMM (W-ADMM), which reduces communication costs at the expense of running time. To accelerate the convergence speed of W-ADMM, we propose the parallel random walk ADMM (PW-ADMM) algorithm, where multiple random walks are active at the same time. Moreover, to further reduce the running time of PW-ADMM, the intelligent parallel random walk ADMM (IPW-ADMM) algorithm is proposed through integrating the *Random Walk with Choice* with PW-ADMM. By numerical results from simulations, we demonstrate that the proposed algorithms can be both communication efficient and fast in running speed compared with state-of-the-art methods.

Index Terms—Decentralized network, consensus optimization, alternating direction method of multipliers (ADMM).

I. INTRODUCTION

CONSIDER a network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{1, \dots, N\}$ is the set of agents and \mathcal{E} is the set of links. The agents aim to solve the following consensus optimization problem,

$$\min_x \sum_{i=1}^N f_i(x) \quad (1)$$

where $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$ is the local loss function held by agent i , and all the agents share a common optimization variable $x \in \mathbb{R}^n$. This consensus problem is applied in various areas including wireless sensor networks [1], [2] and smart grid implementations [3]. Specifically in [4], the distributed beam-forming scheme for multiple relay nodes (RNs) is designed by solving a consensus problem. In general [5], consensus-based distributed linear estimation for cooperative communication in wireless networks can be formulated as (1).

A few decentralized algorithms have been provided to solve the consensus problem in (1). For low computa-

tion complexity, the first-order algorithms such as decentralized gradient decent (DGD) and EXTRA are proposed by [6] and [7], respectively, where agents use their local gradient during the optimization process. Among existing decentralized algorithms, the alternating direction method of multipliers (ADMM), which is extensively applied in wireless communications [8], is shown to be faster than DGD in convergence [9], in which at every iteration an agent needs to solve an optimization problem with collected information from neighboring agents. Besides, a variety of algorithms such as Gauss-Seidel ADMM and Jacobi-Proximal ADMM [10] based on the original work in [11] are provided to solve the consensus problem (1).

In practice, one ideal decentralized approach is expected to obtain the optimal solution of (1) with the minimal communication and computation costs. Lots of research efforts have been put on computation complexity reduction. But there are very few results [12]–[14] on reducing the communication cost of ADMM. Though both proposed algorithms, distributed ADMM (D-ADMM) in [12] and communication-censored ADMM (COCA) in [13], limit the overall communication at each iteration, the COCA can adaptively determine whether a message is informative, and D-ADMM relies more on the network topology. The random walk ADMM (W-ADMM) algorithm is proposed in [14], which randomly activates a succession of nodes and incrementally updates the optimization variable. W-ADMM can achieve much less communication cost but at the expense of running time, since at each iteration only one agent is active for optimization. However, all the approaches provided in [12]–[14] are synchronous ADMM, which may suffer from the straggler problem [15].

In what follows, we will propose the parallel random walk ADMM (PW-ADMM) algorithm that allows multiple random walks active in parallel. Furthermore, we integrate the intelligent agents selection scheme with PW-ADMM, which is presented in the algorithm of intelligent parallel random walk ADMM (IPW-ADMM). By numerical results, we show that the proposed approaches can be both communication efficient and fast in running time.

The remaining of this letter is organized as follows. We first introduce the parallel random walk algorithms in Section II. To demonstrate the effectiveness of the proposed approaches, we provide numerical results in Section III. Finally, we conclude the letter in Section IV.

II. PARALLEL RANDOM WALK ALGORITHMS

By defining $\mathbf{x} = [x_1, \dots, x_N] \in \mathbb{R}^{nN}$, problem (1) can be rewritten as

$$\min_{\mathbf{x}, \mathbf{z}} \sum_{i=1}^N f_i(x_i), \quad \text{s.t. } \mathbb{1} \otimes \mathbf{z} - \mathbf{x} = \mathbf{0}, \quad (2)$$

Manuscript received October 1, 2019; revised November 11, 2019; accepted November 16, 2019. Date of publication November 25, 2019; date of current version February 11, 2020. The work of Ming Xiao was supported in part by the EU Marie Skłodowska-Curie Actions Project entitled “High-reliability Low-latency Communications with network coding”, and the ERA-NET Smart Energy Systems SG+ 2017 Program, “SMART-MLA” with Project number 89029 (and SWEA number 42811-2). The work of Zheng Ma was supported in part by the National Natural Science Foundation of China (no. U1734209, no. U1709219), in part by the Key Project of Sichuan Province (no. 2017SZYZF0002), in part by the Key International Cooperation Project of Sichuan Province (no. 2017HH0002), in part by the Marie Curie Fellowship (no. 796426), and in part by the NSFC China-Swedish project (no. 6161101297) and 111 project (No.111-2-14). The associate editor coordinating the review of this letter and approving it for publication was G. C. Alexandropoulos. (*Corresponding author: Yu Ye.*)

The authors are with the School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, 114 28 Stockholm, Sweden (e-mail: yu9@kth.se; haoch@kth.se; zma@kth.se, mingx@kth.se).

Digital Object Identifier 10.1109/LCOMM.2019.2955442

1558-2558 © 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.

See <https://www.ieee.org/publications/rights/index.html> for more information.

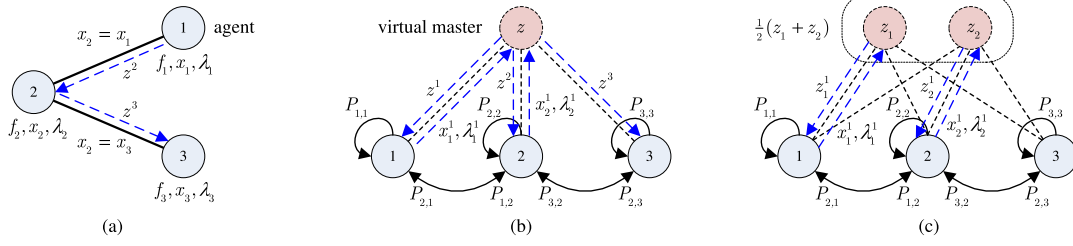


Fig. 1. (a) An example of W-ADMM; (b) the equivalent architecture of W-ADMM; (c) the equivalent architecture of parallel random walk ADMM algorithms.

Algorithm 1 Parallel Walk ADMM (PW-ADMM)

- 1: **Initialize:** $\{x_i^0 = \mathbf{0}, \lambda_i^0 = \mathbf{0}, k_i = 0 | i \in \mathcal{V}\}$ and $\{z_m^1 = \mathbf{0}, k'_m = 0, i^{k'_m} | i^{k'_m} \in \mathcal{V}, i^{k'_m} \neq i^{k'_l}, \forall l \neq m, l, m \in \mathcal{M}\}$
- 2: **Algorithm of the m -th Walk:**
- 3: **for** $k'_m = 0, 1, 2, \dots$ **do**
- 4: update $x_i^{k'_m+1}$ according to (9a) with $i = i^{k'_m}$;
- 5: update $\lambda_i^{k'_m+1}$ according to (9b) with $i = i^{k'_m}$;
- 6: update $z_m^{k'_m+2}$ according to (9c) with $i = i^{k'_m}$;
- 7: set $k_i \leftarrow k_i + 1$ with $i = i^{k'_m}$;
- 8: choose $i^{k'_m+1} (\in \bar{\mathcal{V}}_{i^{k'_m}})$ according to $P_{i^{k'_m}, i^{k'_m+1}}$;
- 9: send $z_m^{k'_m+2}$ to node $i^{k'_m+1}$.
- 10: **end for**

where $z \in \mathbb{R}^n$, $\mathbf{1} = [1, \dots, 1]^T \in \mathbb{R}^n$ and \otimes is Kronecker product. The augmented Lagrangian for problem (2) is

$$\mathcal{L}_\rho(\mathbf{x}, \mathbf{z}, \boldsymbol{\lambda}) = \sum_{i=1}^N f_i(x_i) + \langle \boldsymbol{\lambda}, \mathbf{1} \otimes \mathbf{z} - \mathbf{x} \rangle + \frac{\rho}{2} \|\mathbf{1} \otimes \mathbf{z} - \mathbf{x}\|^2, \quad (3)$$

where $\boldsymbol{\lambda} = [\lambda_1, \dots, \lambda_N] \in \mathbb{R}^{nN}$ is the dual variable, and $\rho > 0$ is a constant parameter. The iterated updates of \mathbf{x} , $\boldsymbol{\lambda}$ and \mathbf{z} can be found in Algorithm 1 (W-ADMM) in [14]. The Fig. 1 (a) presents an example of W-ADMM. Ignoring the difference in communication cost, the equivalent implementation of W-ADMM is shown in Fig. 1 (b), where agent i updates local variables x_i and λ_i after receiving token z , while the virtual master updates z with up-to-date x_i and λ_i .

A. Parallel Random Walk ADMM

To introduce the parallel random walk ADMM, we extend the architecture in Fig. 1 (b) to multiple virtual masters $\mathcal{M} = \{1, \dots, M\}$ as Fig. 1 (c). Denoting $\mathbf{z} = [z_1, \dots, z_M] \in \mathbb{R}^{nM}$, the decentralized problem is given by

$$\min_{\mathbf{x}, \mathbf{z}} \sum_{i=1}^N f_i(x_i), \quad \text{s.t. } \mathbf{1} \otimes \frac{1}{M} \sum_{l=1}^M z_l - \mathbf{x} = \mathbf{0} \quad (4)$$

where $z_l \in \mathbb{R}^n$ ($l \in \mathcal{M}$) is the token held by the l -th random walk. In the constraint of problem (4), we let x_i equal to the average of the summation of tokens. By doing this, we will be able to update tokens of multiple random walks in parallel. The augmented Lagrangian for problem (4) is

$$\mathcal{L}_\rho(\mathbf{x}, \mathbf{z}, \boldsymbol{\lambda}) = \sum_{i=1}^N f_i(x_i) + \langle \boldsymbol{\lambda}, \mathbf{1} \otimes \frac{1}{M} \sum_{l=1}^M z_l - \mathbf{x} \rangle + \frac{\rho}{2} \|\mathbf{1} \otimes \frac{1}{M} \sum_{l=1}^M z_l - \mathbf{x}\|^2. \quad (5)$$

Following the traditional synchronous ADMM [11], the update for $k+1$ -th iteration follows

$$\frac{1}{M} \sum_{l=1}^M z_l^{k+1} := \frac{1}{N} \sum_{i=1}^N (x_i^k - \frac{\lambda_i^k}{\rho}); \quad (6a)$$

$$x_i^{k+1} := \arg \min_{x_i} f_i(x_i) + \langle \lambda_i^k, \frac{1}{M} \sum_{l=1}^M z_l^{k+1} - x_i \rangle + \frac{\rho}{2} \left\| \frac{1}{M} \sum_{l=1}^M z_l^{k+1} - x_i \right\|^2, \quad i \in \mathcal{V}; \quad (6b)$$

$$\lambda_i^{k+1} := \lambda_i^k + \rho \left(\frac{1}{M} \sum_{l=1}^M z_l^{k+1} - x_i^{k+1} \right), \quad i \in \mathcal{V}. \quad (6c)$$

Inspired by the incremental update of WADMM [14], we transform (6a)-(6c) to the following process by approximating $\frac{1}{M} \sum_{l=1}^M z_l$ with z_m for the update of x_i and λ_i .

$$x_i^{k+1} := \begin{cases} \arg \min_{x_i} f_i(x_i) + \langle \lambda_i^k, z_m^{k+1} - x_i \rangle + \frac{\rho}{2} \|z_m^{k+1} - x_i\|^2 \\ \quad + \frac{1}{2} \|x_i - x_i^k\|_{q_i}^2, & i = i_m, m \in \widetilde{\mathcal{M}}_{k+1}; \\ x_i^k, & \text{o.w.}; \end{cases}$$

$$\lambda_i^{k+1} := \begin{cases} \lambda_i^k + \rho (z_m^{k+1} - x_i^{k+1}), & i = i_m, m \in \widetilde{\mathcal{M}}_{k+1}; \\ \lambda_i^k, & \text{o.w.}; \end{cases} \quad (7a)$$

$$\frac{1}{M} \sum_{l=1}^M z_l^{k+2} := \frac{1}{M} \sum_{l=1}^M z_l^{k+1} + \sum_{i=1}^N \left[\frac{1}{N} (x_i^{k+1} - \frac{\lambda_i^{k+1}}{\rho}) - \frac{1}{N} (x_i^k - \frac{\lambda_i^k}{\rho}) \right], \quad (7b)$$

where $\widetilde{\mathcal{M}}_{k+1} \subseteq \mathcal{M}$ is the set of active random walks at iteration k . We adopt proximal update for x_i with $q_i = \tau_i I$, where $\tau_i > 0$ is a step size penalty chosen by agent i and $\|u\|_G^2 = u^T G u$ is G -norm. Note $i_m \neq i_l$ ($l, m \in \widetilde{\mathcal{M}}_{k+1}, l \neq m$) in (7a) and (7a). Hence to make (7b) satisfied, the update for token z_m should follow

$$z_m^{k+2} := \begin{cases} z_m^{k+1} + \frac{M}{N} (x_{i_m}^{k+1} - \frac{\lambda_{i_m}^{k+1}}{\rho}) - \frac{M}{N} (x_{i_m}^k - \frac{\lambda_{i_m}^k}{\rho}), \\ \min \widetilde{\mathcal{M}}_{k+1}; z_m^{k+1}, & \text{o.w.}. \end{cases} \quad (8)$$

The update of z_m^{k+2} in (8) can be carried out in parallel and asynchronously since it does not require information of

$z_l^{k+1}(l, m \in \widetilde{\mathcal{M}}_{k+1}, l \neq m)$. Thus we parallelize (7a)-(7b) by the following updates for $(k'_m + 1)$ -th step of the m -th random walk,

$$\begin{aligned} x_i^{k_i+1} &:= \arg \min_{x_i} f_i(x_i) + \langle \lambda_i^{k_i}, z_m^{k'_m+1} - x_i \rangle \\ &\quad + \frac{\rho}{2} \|z_m^{k'_m+1} - x_i\|^2 + \frac{1}{2} \|x_i - x_i^{k_i}\|_{q_i}^2, \quad i = i^{k'_m}; \end{aligned} \quad (9a)$$

$$\lambda_i^{k_i+1} := \lambda_i^{k_i} + \rho(z_m^{k'_m+1} - x_i^{k_i+1}), \quad i = i^{k'_m}; \quad (9b)$$

$$\begin{aligned} z_m^{k'_m+2} &:= z_m^{k'_m+1} + \frac{M}{N}(x_i^{k_i+1} - \frac{\lambda_i^{k_i+1}}{\rho}) - \frac{M}{N}(x_i^{k_i} - \frac{\lambda_i^{k_i}}{\rho}), \\ &\quad i = i^{k'_m}, \end{aligned} \quad (9c)$$

where k'_m is the clock held by the m -th random walk. In (9a), For agents $i \neq i^{k'_m}$, the local variables x_i and λ_i are not updated by the m -th random walk. The update of $z_m^{k'_m+2}$ only depends on $x_i^{k_i+1}$ and $\lambda_i^{k_i+1}$ from agent $i = i^{k'_m}$ instead of (6a). Defining $\mathcal{V}_i \subset \mathcal{V}$ the set of neighbors of agent $i \in \mathcal{V}$ and $\widetilde{\mathcal{V}}_i = \mathcal{V}_i \cup i$, we present PW-ADMM in Algorithm 1. Similar to W-ADMM, the transition of token z_m follows the embedded Markov chain with probability matrix $\mathbf{P} \in \mathbb{R}^{N \times N}$. When $M = 1$, the PW-ADMM reduces to W-ADMM.

B. Intelligent Parallel Random Walk ADMM

For general problems, the convergence speeds of W-ADMM and PW-ADMM are mainly determined by how frequently all of the agents are visited. Since the transition of the token is determined by probability matrix \mathbf{P} , it is possible that the variables x_i and λ_i at some agents are updated for much fewer rounds than others. This hence may reduce the overall convergence speed. To guarantee agents not to be inactive for long time, we should improve the transition strategy for tokens. Inspired by [16], which introduces the *Random Walk with Choice*, we present IPW-ADMM in Algorithm 2. Different from PW-ADMM, IPW-ADMM requires that agent i has the knowledge of the active rounds of agents in $\widetilde{\mathcal{V}}_i$. Considering the k'_m -th step of the m -th random walk, the updated z_m will be sent to the least visited agent $i^{k'_m+1} := \arg \min_{i \in \mathcal{V}_{i^{k'_m}}} k_i$. Note that we do not count the communication cost of sharing $\{k_i\}$ across the agents since the cost is negligible compared with transmitting tokens.

Algorithm 2 Intelligent PW-ADMM (IPW-ADMM)

- 1: **Initialize:** same as that of PW-ADMM.
 - 2: **Algorithm of the m -th Walk:**
 - 3: follow steps 3-9 of PW-ADMM but substitute step 9 with the following step:
 - 4: choose $i^{k'_m+1} := \arg \min_{i \in \mathcal{V}_{i^{k'_m}}} k_i$.
-

Since all agents and parallel random walks keep individual clock, both PW-ADMM and IPW-ADMM are asynchronous algorithms. However, our proposed algorithms are different from existing work [15], [17], where only one master updates the variable z . Moreover, the updated z is only sent to the agents just active.

C. Convergence Analysis

We present some results on the convergence of PW-ADMM and IPW-ADMM with the following assumption.

Assumption 1: The objective function $f_i(x)$ is L -Lipschitz differentiable, that is

$$\|\nabla f_i(x_1) - \nabla f_i(x_2)\| \leq L \|x_1 - x_2\|, \quad \forall x_1, x_2 \in \mathbb{R}^n. \quad (10)$$

Though (I)PW-ADMM is asynchronous algorithm, we prove the convergence from the synchronous point of view. Without loss of generality, we denote each synchronous iteration as the update for only one token, where $|\widetilde{\mathcal{M}}_k| = 1$.

Theorem 1: Under Assumption 1 and $\tau_i = 0$, the sequence $(x^k, \lambda^k, z^{k+1})$ generated by (I)PW-ADMM satisfies

$$\begin{aligned} &\mathcal{L}_\rho(x^k, \lambda^k, z^{k+1}) - \mathcal{L}_\rho(x^{k+1}, \lambda^{k+1}, z^{k+2}) \\ &\geq \left(\frac{\rho}{2} - \frac{3L}{2} - \frac{L^2}{\rho}\right) \|x_{i_m}^k - x_{i_m}^{k+1}\|^2 \\ &\quad - \frac{N\rho}{2} \left\| \frac{1}{M} \sum_{l=1}^M z_l^{k+1} - z_m^{k+1} \right\|^2, \quad \widetilde{\mathcal{M}}_{k+1} = \{m\}. \end{aligned} \quad (11)$$

Proof: By substituting z^{k+1} with $\frac{1}{M} \sum_{l=1}^M z_l^{k+1}$ in the proof of Lemma 1 and Theorem 1 [14], the result (11) can be obtained. \square

III. NUMERICAL RESULTS

In this section, we provide numerical results from simulations to demonstrate the communication efficiency and running speed of PW-ADMM and IPW-ADMM compared with state-of-the-art methods in [6], [7], [9], [13], [14], [18] with respect to the accuracy, which is defined as

$$\text{accuracy} = \frac{1}{N} \sum_{i=1}^N \frac{\|x_i^k - x^*\|}{\|x_i^0 - x^*\|}, \quad (12)$$

where $x^* \in \mathbb{R}^n$ is optimal solution of (2). For fair comparison, the parameters for algorithms are tuned to be the best, and kept the same in different experiments. The connected network \mathcal{G} is generated randomly with N agents and $|\mathcal{E}| = \frac{N(N-1)}{2}\eta$ links. Besides, the dimension of x_i is set to be $n = 2$. We consider unicast among agents, and the resultant communication cost for each transmission of a n -dimensional vector is 1 unit. The running time includes both computing time and communication time. Without loss of generality, we assume that each agent has multi-process capability to update the tokens for multiple random walks in PW-ADMM and IPW-ADMM. Moreover, the consumed time for each communication is assumed to follow $\mathcal{U}(10^{-5}, 10^{-4})$ (s). The simulation is carried out on a laptop with Intel I7 processor and 8GB memory. The programming environment is Matlab R2016a.

A. Decentralized Least Square Problem

The decentralized least square problem such as [19] aims at solving problem (1) with the local cost function

$$f_i(x_i) = \frac{1}{b_i} \sum_{j=1}^{b_i} \|x_i^T o_{i,j} - t_{i,j}\|^2, \quad (13)$$

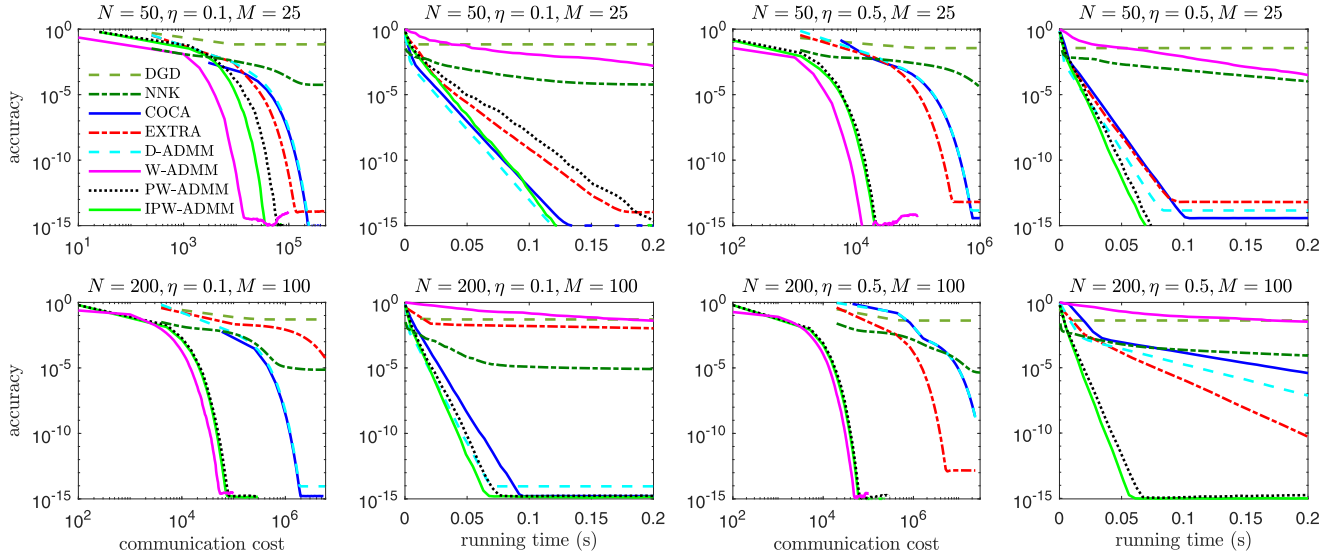


Fig. 2. The accuracy of *decentralized least square* for algorithms: W-ADMM ($\beta = 3$), PW-ADMM ($\rho = 3, \tau = 1.5$), IPW-ADMM ($\rho = 3, \tau = 1.5$), D-ADMM ($\rho = 1$), EXTRA ($\alpha = 0.05$), COCA ($c = 1, \alpha = 1, \rho = 0.85$), DGD ($\alpha = 0.01$), NNK ($K = 2, \alpha = 10^{-3}$).

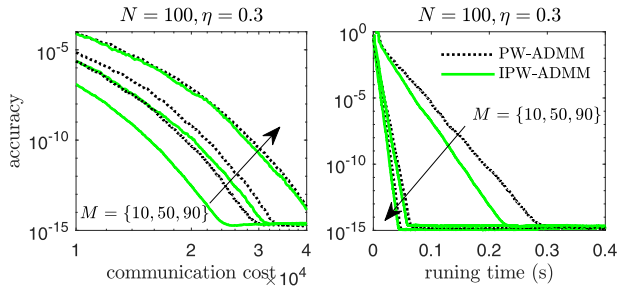


Fig. 3. The impact of M for algorithms: PW-ADMM ($\rho = 1, \tau = 3$), IPW-ADMM ($\rho = 1, \tau = 3$).

where $\mathcal{D}_i = \{o_{i,j}, t_{i,j} | j = 1, \dots, b_i\}$ is the dataset of agent i locally. The entries of input $o_{i,j} \in \mathbb{R}^2$ and target $t_{i,j} \in \mathbb{R}$ follow i.i.d. distribution $\mathcal{U}(0, 1)$. The number of data samples is kept unique across agents with $b_i = 30$.

The accuracy in (12) over communication cost and running time for different network settings is shown in Fig. 2. It is clear that W-ADMM is the most efficient in communication cost but with slow running speed. The proposed parallel random walk algorithms PW-ADMM and IPW-ADMM can significantly reduce the running time from W-ADMM, and consume much less communication resources compared to DGD, D-ADMM, EXTRA and COCA. In addition, IPW-ADMM can further reduce running time from PW-ADMM. Especially when the network is large and highly-connected, i.e., $N = 200$ and $\eta = 0.5$, the PW-ADMM and IPW-ADMM can achieve the best performance in running time with almost the same communication cost with W-ADMM. This is because the inherent asynchronous mechanism of PW-ADMM and IPW-ADMM outperforms the synchronous methods.

In Fig. 3 we present the impact of M , the number of active random walks, on the convergence behavior. It can be concluded that with increasing M , a larger communication cost

of both PW-ADMM and IPW-ADMM is required to achieve the same accuracy, while the running time will be shortened. Hence there exists a trade-off between the communication cost and the running time over M . Besides, for a larger M , e.g. $M = 90$, the accuracy gap between PW-ADMM and IPW-ADMM shrinks compared with the case where $M = 10$. This shows the advantage of intelligently choosing updating path for each walk according to the updated frequency of agents over randomly processing is weakened when more random walks are active.

B. Decentralized Logistic Regression Problem

In the decentralized logistic regression, the local loss function of agent i is

$$f_i(x_i) = \frac{1}{b_i} \sum_{j=1}^{b_i} \log(1 + \exp(-t_{i,j} x_i^T o_{i,j})) \quad (14)$$

where $t_{i,j} \in \{-1, 1\}$ and $b_i = 30$. Each sample feature $o_{i,j}$ follows $\mathcal{N}(0, I)$. To generate $t_{i,j}$, we first generate a random vector $x_0 \in \mathbb{R}^2 \sim \mathcal{N}(0, I)$. Then for each sample, we generate $v_{i,j}$ according to $\mathcal{U}(0, 1)$, and if $v_{i,j} \leq (1 + \exp(-x_0^T o_{i,j}))^{-1}$, we set $t_{i,j}$ as 1, otherwise -1 . Since it is difficult to solve the optimization problem, e.g. (9a), in PW-ADMM, we alternatively use the first-order approximation as

$$f_i(x_i) \approx f_i(x_i^k) + \nabla f_i(x_i^k)(x_i - x_i^k). \quad (15)$$

Fairly, we adopt the first-order approximation for algorithms IPW-ADMM, W-ADMM, D-ADMM and COCA. Fig. 4 presents the accuracy over communication cost and running time. Apparently compared to other benchmarks in [6], [7], [9], [13], [14], only the proposed parallel random walk algorithms PW-ADMM and IPW-ADMM can guarantee both the communication-efficiency and fast convergence speed for different network settings. The curves with different network setups present the similar trends as those of Fig. 2.

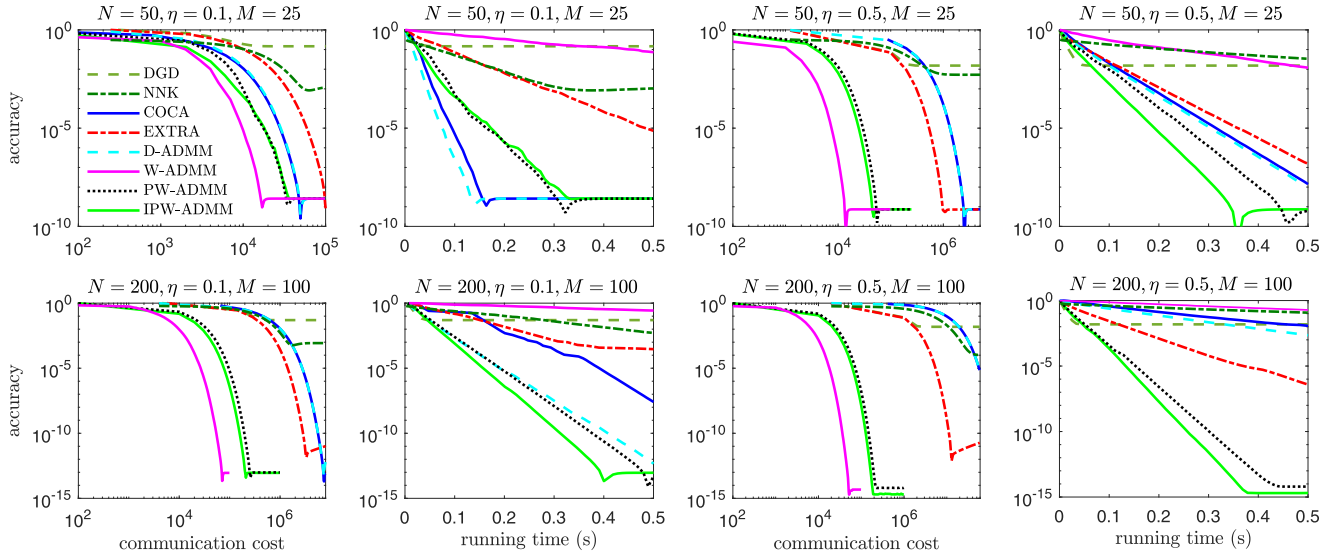


Fig. 4. The accuracy of *decentralized logistic regression* for algorithms: W-ADMM ($\beta = 1$), PW-ADMM ($\rho = 1, \tau = 3$), IPW-ADMM ($\rho = 1, \tau = 3$), D-ADMM ($\rho = 5$), EXTRA ($c = 0.01$), COCA ($c = 5, \alpha = 1, \rho = 0.85$), DGD ($\alpha = 0.01$), NNK ($K = 2, \alpha = 10^{-3}$).

IV. CONCLUSION

We study the communication efficiency and running time for ADMM based consensus problem in decentralized networks. By extending W-ADMM, two parallel random walk algorithms, PW-ADMM and IPW-ADMM are proposed. With the asynchronous characteristic, compared with other approaches, the provided algorithms can achieve much faster running speed with less communication costs, especially for the dense networks. Moreover, simulations demonstrate the scalability of PW-ADMM and IPW-ADMM in terms of the network size.

REFERENCES

- [1] I. D. Schizas, A. Ribeiro, and G. B. Giannakis, "Consensus in ad-hoc WSNs with noisy links—Part I: Distributed estimation of deterministic signals," *IEEE Trans. Signal Process.*, vol. 56, no. 1, pp. 350–364, Jan. 2008.
- [2] Z. Fanni, C. Li, and Z. Tian, "Distributed compressive spectrum sensing in cooperative multihop cognitive networks," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 1, pp. 37–48, Feb. 2011.
- [3] H. J. Liu, W. Shi, and H. Zhu, "Distributed voltage control in distribution networks: Online and robust implementations," *IEEE Trans. Smart Grid*, vol. 9, no. 6, pp. 6106–6117, Nov. 2018.
- [4] J. Choi, "Distributed beamforming using a consensus algorithm for cooperative relay networks," *IEEE Commun. Lett.*, vol. 15, no. 4, pp. 368–370, Apr. 2011.
- [5] H. Paul, J. Fliege, and A. Dekorsy, "In-network-processing: Distributed consensus-based linear estimation," *IEEE Commun. Lett.*, vol. 17, no. 1, pp. 59–62, Jan. 2013.
- [6] 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.
- [7] W. Shi, Q. Ling, G. Wu, and W. Yin, "Extra: An exact first-order algorithm for decentralized consensus optimization," *J. Optim.*, vol. 25, no. 2, pp. 944–966, May 2015.
- [8] E. Vlachos, G. C. Alexandropoulos, and J. Thompson, "Massive MIMO channel estimation for millimeter wave systems via matrix completion," *IEEE Signal Process. Lett.*, vol. 25, no. 11, pp. 1675–1679, Nov. 2018.
- [9] W. Shi, Q. Ling, K. Yuan, G. Wu, and W. Yin, "On the Linear Convergence of the ADMM in Decentralized Consensus Optimization," *IEEE Trans. Signal Process.*, vol. 62, no. 7, pp. 1750–1761, Apr. 2014.
- [10] W. Deng, M.-J. Lai, Z. Peng, and W. Yin, "Parallel multi-block ADMM with $o(1/k)$ convergence," *J. Sci. Comput.*, vol. 71, no. 2, pp. 712–736, May 2017.
- [11] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.
- [12] J. F. C. Mota, J. W. F. Xavier, P. M. Q. 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, May 2013.
- [13] 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, Mar. 2019.
- [14] X. Mao, K. Yuan, Y. Gu, A. H. Sayed, and W. Yin, "Walkman: A communication-efficient random-walk algorithm for decentralized optimization," 2018, *arXiv:1804.06568*. [Online]. Available: <https://arxiv.org/abs/1804.06568>
- [15] R. Zhang and J. Kwok, "Asynchronous distributed ADMM for consensus optimization," in *Proc. Int. Conf. Mach. Learn.*, 2014, pp. 1701–1709.
- [16] C. Avin and B. Krishnamachari, "The power of choice in random walks: An empirical study," in *Proc. 9th ACM Int. Symp. Modeling Anal. Simulation Wireless Mobile Syst.*, 2006, pp. 219–228.
- [17] T. Chang, M. Hong, W. Liao, and X. Wang, "Asynchronous distributed ADMM for large-scale optimization—Part I: Algorithm and convergence analysis," *IEEE Trans. Signal Process.*, vol. 64, no. 12, pp. 3118–3130, Jun. 2016.
- [18] A. Mokhtari, Q. Ling, and A. Ribeiro, "Network Newton distributed optimization methods," *IEEE Trans. Signal Process.*, vol. 65, no. 1, pp. 146–161, Jan. 2017.
- [19] C. Huang, L. Liu, C. Yuen, and S. Sun, "Iterative channel estimation using LSE and sparse message passing for mmWave MIMO systems," *IEEE Trans. Signal Process.*, vol. 67, no. 1, pp. 245–259, Jan. 2019.