



Differentially private dual gradient tracking for distributed resource allocation[☆]

Wei Huo^a, Xiaomeng Chen^{a,*}, Lingying Huang^b, Karl Henrik Johansson^c, Ling Shi^{a,d}

^a Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Hong Kong

^b School of Automation, Southeast University, Nanjing 210096, China

^c Division of Decision and Control Systems, KTH Royal Institute of Technology, Stockholm, Sweden

^d Department of Chemical and Biological Engineering, Hong Kong University of Science and Technology, Hong Kong

ARTICLE INFO

Article history:

Received 27 March 2024

Received in revised form 9 February 2025

Accepted 4 July 2025

Available online 26 August 2025

Keywords:

Distributed resource allocation

Differential privacy

Dual problem

Directed graph

ABSTRACT

This paper investigates privacy issues in distributed resource allocation over directed networks, where each agent holds a private cost function and optimizes its decision subject to a global coupling constraint through local interaction with other agents. Conventional methods for resource allocation over directed networks require all agents to transmit their original data to neighbors, which poses the risk of disclosing sensitive and private information. To address this issue, we propose an algorithm called differentially private dual gradient tracking (DP-DGT) for distributed resource allocation, which obfuscates the exchanged messages using independent Laplacian noise. Our algorithm ensures that the agents' decisions converge to a neighborhood of the optimal solution almost surely. Furthermore, without the assumption of bounded gradients, we prove that the cumulative differential privacy loss under the proposed algorithm is finite even when the number of iterations goes to infinity. To the best of our knowledge, we are the first to simultaneously achieve these two goals in distributed resource allocation problems over directed networks. Finally, numerical simulations on economic dispatch problems within the IEEE 14-bus system illustrate the effectiveness of our proposed algorithm.

© 2025 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

1. Introduction

Resource allocation (RA) is a key issue in fields like smart grids (Yang et al., 2016) and wireless sensor networks (Xiao et al., 2004), where agents collaboratively optimize their objectives while meeting both local and global constraints. Centralized methods, however, face challenges such as single point failures, high communication demands, and high computational costs (Wood et al., 2013), leading to the development of distributed frameworks that rely on agent interactions. The main challenge in distributed resource allocation (DRA) is managing global resource constraints that link all agents' decisions. Zhang

and Chow (2012) introduced a leader–follower consensus algorithm for quadratic objectives. To address the limitation of quadratic cost functions, Yi et al. (2016) proposed a primal–dual algorithm. These works focused on undirected networks with doubly stochastic information mixing matrices. However, bidirectional information flows may incur unnecessary communication costs or may not exist due to sensor power heterogeneity. To tackle this, Yang et al. (2016) introduced a distributed algorithm for unbalanced directed networks using the gradient push-sum method, though it required agents to calculate the eigenvector of the weight matrix asymptotically. Later, Zhang et al. (2020) explored the dual relationship between DRA and distributed optimization (DO), employing the push–pull technique for determining explicit convergence rates.

The existing literature assumes secure transmission of raw data, but the distributed nature of cyber–physical systems raises privacy concerns. Messages between agents are vulnerable to interception, risking theft or inference of sensitive data. For example, in smart grid economic dispatch problems, transmitted messages could expose private user patterns or financial information. Thus, privacy-preserving algorithms for distributed resource allocation are essential. While encryption prevents eavesdropping (Freris & Patrinos, 2016), it may not be feasible for large-scale distributed systems with limited battery power. Beaudé

[☆] The work by W. Huo, X. Chen and L. Shi is supported by the Hong Kong RGC General Research Fund 16211622. The work by L. Huang is supported by the Southeast University project 4008002501. The work by K. H. Johansson was supported in part by Swedish Research Council, Sweden Distinguished Professor Grant 2017-01078, Knut and Alice Wallenberg Foundation, Sweden Wallenberg Scholar Grant, and Swedish Strategic Research Foundation FUSS SUCCESS Grant, Sweden. The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Giacomo Como under the direction of Editor Christos G. Cassandras.

* Corresponding author.

E-mail addresses: whuoaa@connect.ust.hk (W. Huo), xchendu@connect.ust.hk (X. Chen), lingyinghuang@seu.edu.cn (L. Huang), kallej@kth.se (K.H. Johansson), eesling@ust.hk (L. Shi).

et al. (2020) used secure multi-party computation for privacy protection in DRA, but this does not defend against external attackers who intercept all messages. It is also inefficient due to the complexity of computation and communication. Lü et al. (2023) applied conditional noise to protect DRA privacy, but their analysis was limited to quadratic cost functions and only defined privacy as preventing the inference of the exact cost function, which is too narrow and does not cover all scenarios. In contrast, differential privacy (DP) has gained attention for its strong mathematical guarantees and robustness under post-processing (Dwork, 2006).

Recent studies have developed differentially private algorithms for unconstrained distributed optimization (DO) over directed networks. Chen et al. (2023) used state-decomposition and constant noise for privacy-preserving DO, but this only preserves privacy per iteration, leading to cumulative privacy loss over time. Their convergence and privacy analysis also relied on gradient boundedness, which is impractical for applications like distributed economic dispatch with quadratic cost functions. Wang and Nedić (2023) relaxed the gradient boundedness assumption, ensuring ϵ -DP over infinite iterations, but required adjacency gradients to be identical near the optimal point. Huang et al. (2024) showed that gradient tracking-based algorithms cannot achieve ϵ -DP under Laplacian noise with non-summable step sizes. These studies assume strong convexity or convexity of the cost functions. For non-convex objectives, privacy-preserving algorithms have been proposed by Wang and Başar (2023a), Wang and Poor (2022), but they rely on bounded gradient assumptions and apply only to undirected graphs. For DRA with global coupling constraints, Han et al. (2016) achieved DP by adding noise to public signals, but this requires an additional entity for information collection and broadcasting. Ding et al. (2021a) preserved the privacy of cost functions for DRA, though their work is limited to undirected networks. Wang and Nedić (2024) used primal-dual algorithms for finite DP in DRA, but their approach also applies only to undirected graphs and requires additional variables and noise for directed graphs. For directed networks, Lü et al. (2023) employed conditional noise for privacy protection in distributed economic dispatch problems. However, their privacy analysis is limited to quadratic cost functions, and their definition of privacy lacks flexibility and generality. To date, no work addresses differentially private DRA over directed networks.

Motivated by the observations above, our work focuses on providing DP for DRA over directed graphs. We specifically target δ -adjacent DRA problems (Definition 2), which relaxes the bounded gradient assumption. To address global coupling constraints, we consider the dual formulation of DRA. While the dual relationship between DO and DRA, along with existing privacy-preserving DO algorithms for directed networks, is well-established, it is important to note that the dual objective function in DRA is not always strongly convex. As a result, the analysis methods in Pu (2020) and Huang et al. (2024) are not directly applicable. We thus analyze the convergence of gradient-tracking with noisy shared information for *non-convex* objectives, significantly extending the analysis in previous works (Chen et al., 2023; Huang et al., 2024; Pu, 2020; Wang & Nedić, 2023). We derive conditions on the step size and noise to ensure both convergence and ϵ -DP over infinite iterations. A comparison of key related works is presented in Table 1. In summary, our main contributions are as follows:

(1) We propose a differentially private dual gradient tracking algorithm, abbreviated as DP-DGT (Algorithm 1), to address privacy issues in DRA over directed networks. Our algorithm masks the transmitted messages in networks with Laplacian noise and does not rely on any extra central authority.

- (2) With the derived sufficient conditions, we prove that the DP-DGT algorithm converges to a neighborhood of the optimal solution (Theorem 2) by showing the convergence of the dual variable even for non-convex objectives (Theorem 1). This theoretical analysis nontrivially extends existing works on gradient tracking with information-sharing noise for convex or strongly convex objectives (Chen et al., 2023; Huang et al., 2024; Pu, 2020; Wang & Nedić, 2023).
- (3) We specify the mathematical expression of privacy loss ϵ under the DP-DGT (Theorem 3) and demonstrate that the DP-DGT algorithm preserves DP for each individual agent's cost function even over infinite iterations (Corollary 2). To our best knowledge, previous studies have only reported differential privacy results for DRA in undirected networks (Ding et al., 2021a). Moreover, our analysis relaxes the gradient assumption used in Chen et al. (2023), Wang and Nedić (2023).

The remainder of this paper is structured as follows: Section 2 introduces the preliminaries and problem formulation for privacy-preserving DRA over directed networks. Section 3 presents a differentially private distributed dual gradient tracking algorithm with robust push-pull. Section 4 details the convergence analysis, followed by a rigorous proof of ϵ -DP over infinite iterations in Section 5. Section 6 provides numerical simulations illustrating the results, and Section 7 discusses conclusions and future research directions.

Notations: Let \mathbb{R}^p and $\mathbb{R}^{p \times q}$ represent the set of p -dimensional vectors and $p \times q$ -dimensional matrices, respectively. The notation $\mathbf{1}_p \in \mathbb{R}^p$ denotes a vector with all elements equal to one, and $I_p \in \mathbb{R}^{p \times p}$ represents a $p \times p$ -dimensional identity matrix. We use $\|\cdot\|_2$ or $\|\cdot\|$ to denote the ℓ_2 -norm of vectors and the induced 2-norm for matrices. We use $\mathbb{P}(\mathcal{A})$ to represent the probability of an event \mathcal{A} , and $\mathbb{E}[x]$ to be the expected value of a random variable x . The notation $\text{Lap}(\theta)$ denotes the Laplace distribution with probability density function $f_L(x|\theta) = \frac{1}{2\theta} e^{-\frac{|x|}{\theta}}$, where $\theta > 0$. If $x \sim \text{Lap}(\theta)$, we have $\mathbb{E}[x^2] = 2\theta^2$ and $\mathbb{E}[x] = 0$.

Graph Theory: A directed graph is denoted as $\mathcal{G} = (\mathcal{N}, \mathcal{E})$, where $\mathcal{N} = \{1, 2, \dots, N\}$ is the set of nodes and $\mathcal{E} \subseteq \mathcal{N} \times \mathcal{N}$ is the edge set consisting of ordered pairs of nodes. Given a nonnegative matrix $\mathbf{A} = [a_{ij}] \in \mathbb{R}^{N \times N}$, the directed graph induced by \mathbf{A} is referred to as $\mathcal{G}_{\mathbf{A}} = (\mathcal{N}, \mathcal{E}_{\mathbf{A}})$, where the directed edge (i, j) from node j to node i exists, i.e., $(i, j) \in \mathcal{E}_{\mathbf{A}}$ if and only if $a_{ij} > 0$. For a node $i \in \mathcal{N}$, its in-neighbor set $\mathcal{N}_{\mathbf{A}, i}^{\text{in}}$ is defined as the collection of all individual nodes from which i can actively and reliably pull data in graph $\mathcal{G}_{\mathbf{A}}$. Similarly, its out-neighbor set $\mathcal{N}_{\mathbf{A}, i}^{\text{out}}$ is defined as the collection of all individual agents that can passively and reliably receive data from node i .

2. Preliminaries and problem statement

This section provides the preliminaries and problem formulations. We first introduce the RA problem over networks and its dual counterpart. Next, we outline the class of algorithms considered and the messages exchanged. We then discuss potential privacy concerns in traditional algorithms and introduce concepts related to DP. Finally, we formulate the problems addressed in this work.

2.1. Resource allocations over networks

We consider a network of N agents that interact on a directed graph \mathcal{G} to collaboratively address a RA problem. Each agent i possesses a local private cost function $F_i : \mathbb{R}^m \rightarrow \mathbb{R}$. Their

Table 1
A comparison of some related Decentralized Algorithms.

	Problem	Topology	Gradient assumption	DP consideration
Zhang et al. (2020)	DRA	Directed	No assumption	×
Ding et al. (2021a)	DRA	Undirected	Same adjacent gradients in the horizontal position	ϵ -DP over infinite iterations
Wang and Poor (2022)	DO (non-convex)	Undirected	Uniformly bounded gradient	Information-theoretic privacy at each iteration
Wang and Başar (2023a)	DO (non-convex)	Undirected	Uniformly bounded gradient	(ϵ, δ) -DP at each iteration
Chen et al. (2023)	DO (strongly convex)	Directed	Uniformly bounded gradient	ϵ -DP over finite iterations
Wang and Nedić (2023)	DO (convex)	Directed	Same adjacent gradients near the optimal point	ϵ -DP over infinite iterations
Huang et al. (2024)	DO (strongly convex)	Directed	Bounded distance between adjacent gradients	ϵ -DP over infinite iterations
Our work	DRA	Directed	Bounded distance between adjacent gradients	ϵ -DP over infinite iterations

goal is to solve the following resource allocation problem using a distributed algorithm over \mathcal{G} :

$$\begin{aligned} \min_{\mathbf{w} \in \mathbb{R}^{N \times m}} \quad & F(\mathbf{w}) = \sum_{i=1}^N F_i(w_i), \\ \text{s.t.} \quad & \sum_{i=1}^N w_i = \sum_{i=1}^N d_i, \quad w_i \in \mathcal{W}_i, \quad \forall i \in \mathcal{N}, \end{aligned} \quad (1)$$

where $w_i \in \mathbb{R}^m$ represents the local decision of agent i , indicating the resource allocated to the agent, $\mathcal{W}_i \subseteq \mathbb{R}^m$ refers to the local closed and convex constraint set, $\mathbf{w} = [w_1, \dots, w_N]^T \in \mathbb{R}^{N \times m}$, and d_i denotes the local private resource demand of agent i . Let $d = \sum_{i=1}^N d_i$, and thus $\sum_{i=1}^N w_i = d$ represents the overall balance between supply and demand, indicating the coupling among agents.

Throughout the paper, we make the following assumptions:

Assumption 1 (Strong Convexity and Slater's Condition).

- (1) The local cost function F_i is μ -strongly convex for all $i \in \mathcal{N}$, i.e., for any $w, w' \in \mathbb{R}^m$, $\|\nabla F_i(w) - \nabla F_i(w')\| \geq \mu \|w - w'\|$.
- (2) There exists at least one point in the relative interior \mathcal{W} that can satisfy the power balance constraint $\sum_{i=1}^N w_i = \sum_{i=1}^N d_i$, where $\mathcal{W} = \mathcal{W}_1 \times \dots \times \mathcal{W}_N$.

Assumption 1 ensures strong duality between problem (1) and its dual counterpart, enabling us to deal with the coupling constraint based on its dual problem.

2.2. Dual problem

To handle the global constraint, we begin by formulating the dual problem of (1). The Lagrange function of (1) is given by

$$\mathcal{L}(\mathbf{w}, x) = \sum_{i=1}^N F_i(w_i) + x^T \left(\sum_{i=1}^N w_i - \sum_{i=1}^N d_i \right), \quad (2)$$

where $x \in \mathbb{R}^m$ represents the dual variable. Thus, the dual problem of (1) can be expressed by

$$\max_{x \in \mathbb{R}^m} \inf_{\mathbf{w} \in \mathcal{W}} \mathcal{L}(\mathbf{w}, x). \quad (3)$$

The objective function in (3) can be written as

$$\begin{aligned} & \inf_{\mathbf{w} \in \mathcal{W}} \mathcal{L}(\mathbf{w}, x) \\ &= \sum_{i=1}^N \inf_{w_i \in \mathcal{W}_i} (F_i(w_i) + x^T w_i) - x^T \sum_{i=1}^N d_i \end{aligned}$$

$$= \sum_{i=1}^N -F_i^*(-x) - x^T \sum_{i=1}^N d_i,$$

where

$$F_i^*(x) = \sup_{w_i \in \mathcal{W}_i} (x^T w_i - F_i(w_i)) \quad (4)$$

corresponds to the convex conjugate function for the pair (F_i, \mathcal{W}_i) (Bertsekas, 1997). Consequently, the dual problem (3) can be formulated as the subsequent DO problem

$$\min_{x \in \mathbb{R}^m} f(x) = \sum_{i=1}^N f_i(x), \quad \text{where } f_i(x) \triangleq F_i^*(-x) + x^T d_i. \quad (5)$$

According to the Fenchel duality between strong convexity and the Lipschitz continuous gradient (Boyd & Vandenberghe, 2004; Zhou, 2018), the strong convexity of F_i indicates the differentiability of F_i^* with Lipschitz continuous gradients, and the supremum in (4) can be attained. Danskin's theorem states that $\nabla F_i^*(x) = \arg \max_{w_i \in \mathcal{W}_i} \{x^T w_i - F_i(w_i)\}$ (Bertsekas, 1997), providing the gradient of F_i^* . Therefore, we have

$$\begin{aligned} \nabla f_i(x) &= -\nabla F_i^*(-x) + d_i \\ &= -\arg \min_{w_i \in \mathcal{W}_i} \{F_i(w_i) + x^T w_i\} + d_i. \end{aligned} \quad (6)$$

We find that the dual gradient $\nabla f_i(x)$ captures the local deviation or mismatch between resource supply and demand, i.e., $w_i - d_i$, in some sense.

If **Assumption 1** holds, we observe the strong duality between the dual DO problem (3) and the primal DRA problem (1). This equivalence is captured through $F^* = -f^*$ and the optimal solution x^* of (3) satisfies $F_i^*(w_i^*) + F_i^*(-x^*) = -x^{*T} w_i^*$. As a result, if the proposed algorithm can drive the dual variable in (5) to the optimal one, it equivalently steers \mathbf{w} to the optimal solution \mathbf{w}^* in (1). Hence, our algorithmic focus can be directed towards solving (5), the standard DO problem.

2.3. Communication networks and information flows

Let us introduce the class of DRA algorithms we are considering in this paper. Gradient-tracking with push-pull (Pu et al., 2020) is one of the DO algorithms that enable agents to solve optimization problems in directed networks, particularly for those unbalanced networks lacking doubly stochastic weight matrices. Solving the dual counterpart (5) using this algorithm allows agents to obtain the optimal solution for the DRA problem (1) over \mathcal{G} . Specifically, agent i maintains a local estimate of the dual variable x and a local estimate of the global constraint deviation at iteration k , denoted as $\tilde{w}_{i,k}$ and $z_{i,k}$, respectively. These two

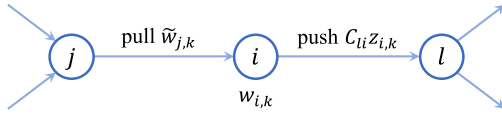


Fig. 1. Information flows of agent i under algorithm \mathcal{A} .

local variables are shared using two different communication networks, \mathcal{G}_R and \mathcal{G}_{C^T} , respectively. These two networks are induced by matrices $\mathbf{R} = [R_{ij}] \in \mathbb{R}^{N \times N}$ and $\mathbf{C} = [C_{ij}] \in \mathbb{R}^{N \times N}$, respectively, where $R_{ij} > 0$ for any $j \in \mathcal{N}_{R,i}^{\text{in}}$ and $C_{ij} > 0$ for any $i \in \mathcal{N}_{C,j}^{\text{out}}$. We call this special class of algorithms *DRA with dual gradient tracking*, denoted as \mathcal{A} . A representative form of \mathcal{A} that we consider in this paper is as follows (Zhang et al., 2020):

$$\tilde{w}_{i,k+1} = \sum_{j=1}^N R_{ij} \tilde{w}_{j,k} + \beta_k z_{i,k}, \quad (7a)$$

$$w_{i,k+1} = \arg \min_{w \in \mathcal{W}_i} \{F_i(w) - \tilde{w}_{i,k+1}^T w\}, \quad (7b)$$

$$z_{i,k+1} = \sum_{j=1}^N C_{ij} z_{j,k} - \iota(w_{i,k+1} - w_{i,k}), \quad (7c)$$

where $\beta_k > 0$ and $\iota > 0$ are the step sizes. Fig. 1 illustrates the information flows under algorithm \mathcal{A} over the communication network. At each iteration, agent i pushes $C_{ii}z_{i,k}$ to each out-neighboring agent $l \in \mathcal{N}_{C,i}^{\text{out}}$, pulls the dual variable estimate $\tilde{w}_{j,k}$ from its in-neighboring agent $j \in \mathcal{N}_{R,i}^{\text{in}}$, and updates its privately-owned primal optimization variable $w_{i,k}$.

We impose the following assumptions on the communication graphs:

Assumption 2. The graphs \mathcal{G}_R and \mathcal{G}_{C^T} each contains at least one spanning tree. Moreover, there exists at least one node that is a root of a spanning tree for both \mathcal{G}_R and \mathcal{G}_{C^T} .

Assumption 3. The matrix \mathbf{R} is row-stochastic and \mathbf{C} is column-stochastic, i.e., $\mathbf{R}\mathbf{1} = \mathbf{1}$ and $\mathbf{1}^T \mathbf{C} = \mathbf{1}^T$.

Assumption 2 is less restrictive than previous works such as Tsianos et al. (2012) and Xi and Khan (2017), as it does not necessitate a strongly connected directed graph. This allows more flexibility in graph design. However, directly transmitting $\tilde{w}_{j,k}$ and $C_{ii}z_{i,k}$ in the network will pose privacy concerns, which we will discuss in the next subsection.

2.4. Differential privacy

In deterministic optimization problems, given a specific initialization and topology, the generated data and decisions are uniquely determined by the cost function of each agent. Therefore, in insecure networks, agents should protect the privacy of their cost functions against eavesdroppers while calculating the optimal solution to (1) in a distributed manner. In this paper, we consider the following commonly used eavesdropping attack model (Chen et al., 2023; Wang & Nedić, 2023):

Definition 1. An eavesdropping attack is an adversary that is able to listen to all communication messages in the network.

Note that our definition of eavesdropping models a powerful attack as the adversary potentially can intercept every message in the network. For example, in the communication network shown in Fig. 1, an adversary under eavesdropping attack can obtain $\{\tilde{w}_{i,k}, C_{ii}z_{i,k} | \forall i \in \mathcal{N}, k \geq 0\}$. With this observation, the attacker

is able to learn $w_{i,k}$ based on publicly known \mathbf{R}, \mathbf{C} , and step sizes. If the local constraint set \mathcal{W}_i is equal to \mathbb{R}^m and F_i is differentiable, the step (7b) can be rewritten as $w_{i,k+1} = \nabla^{-1}F_i(\tilde{w}_{i,k+1})$, where $\nabla^{-1}F_i$ represents the inverse function of ∇F_i such that $\nabla F_i(w_{i,k+1}) = \tilde{w}_{i,k+1}$. As a result, the cost function F_i could be deduced from (7b), which causes privacy leakage.

The potential privacy leakage in algorithms in \mathcal{A} motivates us to design novel privacy-preserving algorithms to protect the privacy of agents' cost functions. To measure privacy, we introduce concepts associated with DP (Dwork, 2006).

First, we formulate the closeness of DRA problems. We denote the DRA problem shown in (1) as \mathcal{P} and represent it by four parameters $(\mathcal{W}, \mathcal{S}, F, \mathcal{G})$, where $\mathcal{S} \subseteq \{\mathbb{R}^m \rightarrow \mathbb{R}\}$ is a set of real-valued cost functions, and $F = \sum_{i=1}^N F_i$ with $F_i \in \mathcal{S}$ for each $i \in \mathcal{N}$. Specifically, we define δ -adjacency of two DRA problems by measuring the distance between gradients of the individual's local cost function.

Definition 2 (δ -adjacency). Two distributed resource allocation problems \mathcal{P} and \mathcal{P}' are δ -adjacent if the following conditions hold:

- (1) they share identical domains for resource allocation and communication graphs, i.e., $\mathcal{W} = \mathcal{W}'$ and $\mathcal{G} = \mathcal{G}'$;
- (2) there exists an $i_0 \in \mathcal{N}$ such that $F_{i_0} \neq F'_{i_0}$, and for all $i \neq i_0$, $F_i = F'_i$;
- (3) the distance of gradients of F_{i_0} and F'_{i_0} are bounded by δ on \mathcal{W}_{i_0} , i.e., $\sup_{w \in \mathcal{W}} \|\nabla F_{i_0}(w) - \nabla F'_{i_0}(w)\| \leq \delta$ for any $w \in \mathcal{W}_{i_0}$.

Remark 1. According to Definition 2, two DRA problems are adjacent when the cost function of a single agent changes, while all other conditions remain unchanged. As shown in Table 1, various works define δ -adjacency differently for DP analysis. Many earlier studies relied on the bounded gradient assumption, but this fails for quadratic cost functions commonly used in resource allocation problems (Kar & Hug, 2012; Lü et al., 2023; Yang et al., 2016). Our assumption relaxes this requirement, allowing for more general cost functions. Additionally, unlike (Wang & Nedić, 2023), we do not require adjacent gradients to be identical, nor do we require cost functions to be the same around the optimal value, as in Wang and Başar (2023a), Wang and Nedić (2024).

Remark 2. In this paper, we focus on adjacent functions with different objective functions but identical coupling constraints, aiming to protect the private information in each agent's local objective function, denoted as f_i . This approach is common in privacy-preserving distributed resource allocation problems, such as those in Ding et al. (2021b) and Hu et al. (2023). Works that address private coupling constraints, such as (Munoz et al., 2021), are beyond the scope of our study.

Traditional algorithms in \mathcal{A} do not have any privacy protection in general (Zhang et al., 2020). To preserve the privacy of agents, we should add some random perturbations or uncertainties to confuse the eavesdropper. We denote the class of algorithms \mathcal{A} with random perturbations as \mathcal{R} . Given a DRA problem \mathcal{P} , an iterative randomized algorithm can be considered as a mapping as $\mathcal{R}_{\mathcal{P}}(v_0) : v_0 \rightarrow \mathcal{O}$, where v_0 is the set of initial states of the algorithm \mathcal{R} and \mathcal{O} is the observation sequence of all shared messages.

Let us now define privacy for such a randomized algorithm following the classical ϵ -DP notion introduced by Dwork (2006).

Definition 3 (ϵ -DP). For a given $\epsilon > 0$, a randomized iterative algorithm \mathcal{R} solving (1) is ϵ -DP if for any two δ -adjacent resource

allocation problems \mathcal{P} and \mathcal{P}' , any set of observation sequences $\mathcal{O} \subseteq \text{Range}(\mathcal{R})^1$ and any initial state v_0 , it holds that

$$\mathbb{P}[\mathcal{R}_{\mathcal{P}}(v_0) \in \mathcal{O}] \leq e^\epsilon \mathbb{P}[\mathcal{R}_{\mathcal{P}'}(v_0) \in \mathcal{O}], \quad (8)$$

where the probability is over the randomness introduced in each iteration of the algorithm.

Definition 3 defines ϵ -differential privacy (DP) for a randomized algorithm $\mathcal{R}(\cdot)$, ensuring minimal differentiation in output probabilities for two δ -adjacent RA problems. Here, ϵ represents the privacy budget or loss, with smaller values making it harder for an eavesdropper to distinguish between two sets of cost functions based on observed data. However, in many existing works (Chen et al., 2022, 2023; Wang & Başar, 2023b), a finite cumulative privacy budget is only achieved over a limited number of iterations. As iterations increase, even if the solution nears optimal, privacy may eventually be compromised. Thus, we carefully design the step size and noise parameters to preserve privacy.

2.5. Problem formulation

This work first designs a novel distributed algorithm \mathcal{R} that preserves privacy for RA problems on directed graphs. We then analyze the conditions on step sizes and perturbations that ensure both convergence and ϵ -DP over infinite iterations.

3. Algorithm development

DP is typically maintained by introducing noise into transmitted data. However, information-sharing noise corrupts exchanged messages, causing agents to receive distorted estimates of dual variables and constraint deviations, which reduces accuracy. This creates a fundamental trade-off between privacy and optimization accuracy. To understand the impact of noise, we start by analyzing the update of $\tilde{w}_{i,k}$ and $z_{i,k}$ under the traditional algorithm (7) in the presence of information-sharing noise.

Defining $\tilde{\mathbf{w}}_k = [\tilde{w}_{1,k}, \dots, \tilde{w}_{N,k}]^T \in \mathbb{R}^{N \times m}$ and $\mathbf{z}_k = [z_{1,k}, \dots, z_{N,k}]^T \in \mathbb{R}^{N \times m}$, we can rewrite (7a) and (7c) in their compact forms:

$$\tilde{\mathbf{w}}_{k+1} = \mathbf{R}\tilde{\mathbf{w}}_k + \beta_k \mathbf{z}_k, \quad (9a)$$

$$\mathbf{z}_{k+1} = \mathbf{C}\mathbf{z}_k - \iota(\mathbf{w}_{k+1} - \mathbf{w}_k). \quad (9b)$$

By setting $\mathbf{1}^T \mathbf{z}_0 = -\iota \left(\sum_{i=1}^N w_{i,0} - \sum_{i=1}^N d_i \right)$, we can deduce using induction that

$$\mathbf{1}^T \mathbf{z}_k = -\iota \left(\sum_{i=1}^N w_{i,k} - \sum_{i=1}^N d_i \right),$$

which implies that the agents can track the global mismatch between resource supply and demand.

However, when exchanged messages are subject to noise, i.e., the received values are $\tilde{w}_{i,k} + \zeta_{i,k}$ and $z_{i,k} + \xi_{i,k}$ instead of $\tilde{w}_{i,k}$ and $z_{i,k}$, respectively, the update of the conventional algorithm (9) becomes

$$\tilde{\mathbf{w}}_{k+1} = \mathbf{R}(\tilde{\mathbf{w}}_k + \boldsymbol{\zeta}_k) + \beta_k \mathbf{z}_k, \quad (10a)$$

$$\mathbf{z}_{k+1} = \mathbf{C}(\mathbf{z}_k + \boldsymbol{\xi}_k) - \iota(\mathbf{w}_{k+1} - \mathbf{w}_k), \quad (10b)$$

where $\boldsymbol{\zeta}_k = [\zeta_{1,k}, \dots, \zeta_{N,k}]^T \in \mathbb{R}^{N \times m}$, $\boldsymbol{\xi}_k = [\xi_{1,k}, \dots, \xi_{N,k}]^T \in \mathbb{R}^{N \times m}$, and $\zeta_{i,k} \in \mathbb{R}^m$ and $\xi_{i,k} \in \mathbb{R}^m$ are injected noises. Using induction, we can deduce that even under $\mathbf{1}^T \mathbf{z}_0 = -\iota \left(\sum_{i=1}^N w_{i,0} - \right.$

¹ $\text{Range}(\mathcal{R})$ denotes the set of all possible observation sequences under the algorithm \mathcal{R} .

Algorithm 1 Differentially Private Dual Gradient Tracking (DP-DGT)

1: Input: Step size sequence $\{\alpha_k\}$, noise sequences $\{\xi_{i,k}\}$ and $\{\zeta_{i,k}\}$ for any $i \in \mathcal{N}$ and $k \geq 0$, and the parameters γ and ϕ .

2: Initialization: $w_{i,0}, s_{i,0}, \tilde{w}_{i,0} \in \mathbb{R}^m$.

3: **for** $k = 1, 2, \dots$, **do**

4: **for each** $i \in \mathcal{N}$,

5: Agent i pushes $C_{li}(s_{i,k} + \xi_{i,k})$ to each agent $l \in \mathcal{N}_{\mathbf{C}_i}^{\text{out}}$.

6: Agent i pulls $\tilde{w}_{j,k} + \zeta_{j,k}$ from each agent $j \in \mathcal{N}_{\mathbf{R}_i}^{\text{in}}$.

7: **for each** $i \in \mathcal{N}$,

$$s_{i,k+1} = (1 - \gamma)s_{i,k} + \gamma \sum_{j=1}^N C_{ij} (s_{j,k} + \xi_{j,k}) - \alpha_k (w_{i,k} - d_i), \quad (12a)$$

$$w_{i,k+1} = \arg \min_{w \in \mathcal{W}_i} \{F_i(w) - \tilde{w}_{i,k+1}^T w\}, \quad (12b)$$

$$\tilde{w}_{i,k+1} = (1 - \phi)\tilde{w}_{i,k} + \phi \sum_{j=1}^N R_{ij} (\tilde{w}_{j,k} + \zeta_{j,k}) + (s_{i,k+1} - s_{i,k}). \quad (12c)$$

8: **end for**

$$\sum_{i=1}^N d_i):$$

$$\mathbf{1}^T \mathbf{z}_k = -\iota \left(\sum_{i=1}^N w_{i,k} - \sum_{i=1}^N d_i \right) + \sum_{l=0}^{k-1} \mathbf{1}^T \boldsymbol{\xi}_l. \quad (11)$$

In the conventional algorithm, information-sharing noise accumulates over iterations, increasing total variance and significantly affecting optimization accuracy. To overcome the limitations of existing dual gradient-tracking algorithms (Ding et al., 2021a; Zhang et al., 2020) and reduce the impact of this noise, we draw inspiration from the robust push-pull method (Pu, 2020). Instead of sharing the direct per-step global deviation estimate $z_{i,k}$, each agent shares the cumulative deviation estimate $s_{i,k}$. Our privacy-preserving method is outlined in Algorithm 1. The Laplace mechanism is a fundamental technique for achieving DP and we thus assume that the noise satisfies Assumption 4. Although Gaussian noise can also be employed, it may require a slight relaxation of the definition of DP (Dwork, 2006).

Assumption 4. The noise $\xi_{i,k}$ and $\zeta_{i,k}$ are independently drawn by agent i from the following zero-mean Laplace distribution,

$$\xi_{i,k} \sim \text{Lap}(\theta_{\xi,k}), \quad \zeta_{i,k} \sim \text{Lap}(\theta_{\zeta,k}),$$

where $\{\theta_{\xi,k}\}$ and $\{\theta_{\zeta,k}\}$ are sequences to be designed.

Defining $\mathbf{s}_k = [s_{1,k}, \dots, s_{N,k}]^T \in \mathbb{R}^{N \times m}$, we can rewrite (12a) and (12c) from Algorithm 1 as follows:

$$\mathbf{s}_{k+1} = (1 - \gamma)\mathbf{s}_k + \gamma\mathbf{C}(\mathbf{s}_k + \boldsymbol{\xi}_k) - \alpha_k(\mathbf{w}_k - \mathbf{d}), \quad (13a)$$

$$\tilde{\mathbf{w}}_{k+1} = (1 - \phi)\tilde{\mathbf{w}}_k + \phi\mathbf{R}(\tilde{\mathbf{w}}_k + \boldsymbol{\zeta}_k) + (\mathbf{s}_{k+1} - \mathbf{s}_k). \quad (13b)$$

In this compact form, we observe that $\mathbf{s}_{k+1} - \mathbf{s}_k$ is fed into the dual variable update and serves as the deviation estimate. This approach prevents the accumulation of information noise on the global mismatch estimate. In fact, using the update rule of \mathbf{s}_k

in (13a), and by letting $\mathbf{z}_k = \mathbf{s}_{k+1} - \mathbf{s}_k$, we obtain:

$$\begin{aligned} \mathbf{1}^T \mathbf{z}_k &= \mathbf{1}^T (\mathbf{s}_{k+1} - \mathbf{s}_k) \\ &= -\gamma \mathbf{1}^T \mathbf{s}_k + \gamma \mathbf{1}^T \mathbf{C} (\mathbf{s}_k + \xi_k) - \alpha_k \left(\sum_{i=1}^N w_{i,k} - \sum_{i=1}^N d_i \right) \\ &= -\alpha_k \left(\sum_{i=1}^N w_{i,k} - \sum_{i=1}^N d_i \right) + \gamma \mathbf{1}^T \xi_k, \end{aligned} \quad (14)$$

regardless of the initial selection of \mathbf{s}_0 , where we used the property $\mathbf{1}^T \mathbf{C} = \mathbf{1}^T$ from Assumption 3. Thus, the proposed algorithm utilizes $\mathbf{s}_{k+1} - \mathbf{s}_k$ to track the global deviation and prevent noise accumulation in the deviation tracking. Furthermore, in contrast to Ding et al. (2021a), Zhang et al. (2020), our algorithm does not have any requirements on the initialization.

4. Convergence analysis

By leveraging strong duality, we establish the convergence of Algorithm 1 by proving the convergence of the dual problem (5) under robust push-pull with information-sharing noise. However, existing convergence results for dual optimization (DO) with robust push-pull require the objective function f_i to be strongly convex and Lipschitz smooth (Chen et al., 2023; Pu, 2020). It is important to note that the dual objective f_i in (5) often loses strong convexity due to the convex conjugate function F_i^* , even when F_i in (1) is strongly convex. Therefore, we extend the convergence of dual variables under robust push-pull to non-convex objectives. Additionally, Huang et al. (2024) show that ϵ -DP cannot be achieved with Laplace noise if step sizes are not summable (i.e., $\sum_{k=0}^{\infty} \alpha_k = \infty$ and $\sup_{k \geq 0} \alpha_k < \infty$). Thus, only summable step sizes ensure meaningful convergence and privacy performance under Laplacian noise.

We first demonstrate that, with the proposed algorithm and summable step sizes, dual variables for non-convex f_i can converge to a neighborhood of a stationary point. Then, we show the convergence of primal variables in problem (1) under DP-DGT.

4.1. Convergence of dual variables

To better illustrate the difference of our convergence analysis with existing works using push-pull-based gradient-tracking methods, we rewrite the proposed algorithm (12) in a typical push-pull form by letting $y_{i,k} = s_{i,k}$. Since $x_i = -\tilde{w}_i$, we have

$$y_{i,k+1} = (1 - \gamma)y_{i,k} + \gamma \sum_{j=1}^N C_{ij}(y_{j,k} + \xi_{j,k}) + \alpha_k \nabla f_i(x_{i,k}), \quad (15a)$$

$$\begin{aligned} x_{i,k+1} &= (1 - \phi)x_{i,k} + \phi \sum_{j=1}^N R_{ij}(x_{j,k} - \zeta_{j,k}) \\ &\quad - (y_{i,k+1} - y_{i,k}). \end{aligned} \quad (15b)$$

Different from (Ding et al., 2021b), Chen et al. (2023), and Pu (2020), the $f_i, \forall i \in \mathcal{N}$ here is not strongly convex. Let $\mathbf{x}_k = -\tilde{\mathbf{w}}_k$ and $\mathbf{y}_k = -\mathbf{s}_k$ and denote $G(\mathbf{x}) = [\nabla f_1(x_{1,k}), \dots, \nabla f_N(x_{N,k})]^T \in \mathbb{R}^{N \times m}$. Defining

$$\mathbf{C}_\gamma := (1 - \gamma)\mathbf{I} + \gamma\mathbf{C},$$

$$\mathbf{R}_\phi := (1 - \phi)\mathbf{I} + \phi\mathbf{R},$$

we can write (15) in the following compact form:

$$\mathbf{y}_{k+1} = \mathbf{C}_\gamma \mathbf{y}_k + \alpha_k G(\mathbf{x}_k) + \gamma \mathbf{C} \xi_k, \quad (16)$$

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{R}_\phi \mathbf{x}_k + \phi \mathbf{R} \zeta_k - (\mathbf{y}_{k+1} - \mathbf{y}_k) \\ &= \mathbf{R}_\phi \mathbf{x}_k - \mathbf{v}_k + \phi \mathbf{R} \zeta_k - \gamma \mathbf{C} \xi_k, \end{aligned} \quad (17)$$

where we denote $\mathbf{v}_k = (\mathbf{C}_\gamma - \mathbf{I})\mathbf{y}_k + \alpha_k G(\mathbf{x}_k)$. It can be verified that \mathbf{R}_ϕ and \mathbf{C}_γ are row-stochastic and column-stochastic, respectively. Under Assumption 2, we have some preliminary lemmas regarding the communication graphs.

Lemma 1 (Horn & Johnson, 2012). Suppose Assumption 2 holds. The matrix \mathbf{R} has a unique unit nonnegative left eigenvector $\pi_{\mathbf{R}}$ w.r.t. eigenvalue 1, i.e., $\pi_{\mathbf{R}}^T \mathbf{R} = \pi_{\mathbf{R}}^T$ and $\pi_{\mathbf{R}}^T \mathbf{1} = 1$. The matrix \mathbf{C} has a unique unit nonnegative right eigenvector $\pi_{\mathbf{C}}$ w.r.t. eigenvalue 1, i.e., $\mathbf{C}\pi_{\mathbf{C}} = \pi_{\mathbf{C}}$ and $\pi_{\mathbf{C}}^T \mathbf{1} = 1$.

Based on Lemma 1 and the definition of \mathbf{R}_ϕ and \mathbf{C}_γ , we can also deduce that $\pi_{\mathbf{R}}^T \mathbf{R}_\phi = \pi_{\mathbf{R}}^T$ and $\mathbf{C}_\gamma \pi_{\mathbf{C}} = \pi_{\mathbf{C}}$.

Lemma 2 (Pu et al., 2020). Suppose Assumption 2 holds. There exist matrix norms $\|\cdot\|_{\mathbf{R}}$ and $\|\cdot\|_{\mathbf{C}}$ such that $\sigma_{\mathbf{R}} := \|\mathbf{R}_\phi - \mathbf{1}\pi_{\mathbf{R}}^T\|_{\mathbf{R}} < 1$ and $\sigma_{\mathbf{C}} := \|\mathbf{C}_\gamma - \pi_{\mathbf{C}}\mathbf{1}^T\|_{\mathbf{C}} < 1$. Furthermore, $\sigma_{\mathbf{R}}$ and $\sigma_{\mathbf{C}}$ can be arbitrarily closed to the spectral radius of $\mathbf{R}_\phi - \mathbf{1}\pi_{\mathbf{R}}^T$ and $\mathbf{C}_\gamma - \pi_{\mathbf{C}}\mathbf{1}^T$.

Note that the norms $\|\cdot\|_{\mathbf{R}}$ and $\|\cdot\|_{\mathbf{C}}$ are only for matrices, which is defined as

$$\|X\|_{\mathbf{R}} = \|\tilde{\mathbf{R}}X\tilde{\mathbf{R}}^{-1}\|_2 \text{ and } \|X\|_{\mathbf{C}} = \|\tilde{\mathbf{C}}^{-1}X\tilde{\mathbf{C}}\|_2$$

for any matrix $X \in \mathbb{R}^{N \times N}$, where $\tilde{\mathbf{R}}$ and $\tilde{\mathbf{C}}$ are some invertible matrices. To facilitate presentation, we slightly abuse the notations and define vectors norm $\|x\|_{\mathbf{R}} = \|\tilde{\mathbf{R}}x\|_2$ and $\|x\|_{\mathbf{C}} = \|\tilde{\mathbf{C}}^{-1}x\|_2$ for any $x \in \mathbb{R}^N$.

Lemma 3 (Pu et al., 2020). There exist constants $\delta_{\mathbf{R},\mathbf{C}}, \delta_{\mathbf{C},\mathbf{R}}$ such that for all \mathbf{x} , we have $\|\cdot\|_{\mathbf{R}} \leq \delta_{\mathbf{R},\mathbf{C}} \|\cdot\|_{\mathbf{C}}$ and $\|\cdot\|_{\mathbf{C}} \leq \delta_{\mathbf{C},\mathbf{R}} \|\cdot\|_{\mathbf{R}}$. Additionally, we can easily obtain $\|\cdot\|_{\mathbf{R}} \leq \|\cdot\|_2$ and $\|\cdot\|_{\mathbf{C}} \leq \|\cdot\|_2$ from the construction of the norm $\|\cdot\|_{\mathbf{R}}$ and $\|\cdot\|_{\mathbf{C}}$.

Denote $\bar{\mathbf{x}}_k = \mathbf{x}_k^T \pi_{\mathbf{R}}, \bar{\mathbf{v}}_k = \mathbf{v}_k^T \pi_{\mathbf{R}}, \hat{\mathbf{v}}_k = \mathbf{v}_k^T \mathbf{1} = \alpha_k G(\mathbf{x}_k)^T \mathbf{1}$, and let \mathcal{F}_k be the σ -algebra generated by $\{\xi_i, \zeta_i\}_{i=0, \dots, k-1}$. With the above norms, we first establish a system of linear inequalities w.r.t. the expectations of $\|\mathbf{x}_{k+1} - \mathbf{1}_N \bar{\mathbf{x}}_{k+1}^T\|_{\mathbf{R}}^2$ and $\|\mathbf{v}_{k+1} - \pi_{\mathbf{C}} \hat{\mathbf{v}}_{k+1}^T\|_{\mathbf{C}}^2$ for the dual algorithm (15).

Lemma 4. Under Assumptions 2–4 and the L -Lipschitz smoothness of $f_i, \forall i \in \mathcal{N}$, we have the following linear system of inequalities

$$\begin{aligned} &\begin{bmatrix} \mathbb{E} \left[\|\mathbf{x}_{k+1} - \mathbf{1}_N \bar{\mathbf{x}}_{k+1}^T\|_{\mathbf{R}}^2 \mid \mathcal{F}_k \right] \\ \mathbb{E} \left[\|\mathbf{v}_{k+1} - \pi_{\mathbf{C}} \hat{\mathbf{v}}_{k+1}^T\|_{\mathbf{C}}^2 \mid \mathcal{F}_k \right] \end{bmatrix} \\ &\preceq \underbrace{\begin{bmatrix} P_{11,k} & P_{12} \\ P_{21,k} & P_{22,k} \end{bmatrix}}_{\mathbf{P}_k} \begin{bmatrix} \|\mathbf{x}_k - \mathbf{1}_N \bar{\mathbf{x}}_k^T\|_{\mathbf{R}}^2 \\ \|\mathbf{v}_k - \pi_{\mathbf{C}} \hat{\mathbf{v}}_k^T\|_{\mathbf{C}}^2 \end{bmatrix} \\ &+ \underbrace{\begin{bmatrix} B_{11} & B_{12} & B_{13} \\ B_{21} & B_{22} & B_{23,k} \end{bmatrix}}_{\mathbf{B}_k} \begin{bmatrix} \alpha_k^2 \mathbb{E} \left[\|\nabla f(\bar{\mathbf{x}}_k)\|_2^2 \mid \mathcal{F}_k \right] \\ \phi^2 \theta_{\xi,k}^2 \\ \gamma^2 \theta_{\zeta,k}^2 \end{bmatrix}, \end{aligned} \quad (18)$$

where the inequality is taken element-wise, and the elements of matrices \mathbf{P}_k and \mathbf{B}_k are given by

$$P_{11,k} = \frac{1 + \sigma_{\mathbf{R}}^2}{2} + \mathbf{a}_1 \alpha_k^2, \quad P_{12} = \mathbf{a}_2,$$

$$P_{21,k} = (\mathbf{a}_3 + \mathbf{a}_4 \alpha_k^2) \max\{\alpha_{k+1}^2, \alpha_k^2\},$$

$$P_{22,k} = \frac{1 + \sigma_{\mathbf{C}}^2}{2} + \mathbf{a}_5 \max\{\alpha_{k+1}^2, \alpha_k^2\},$$

and

$$\begin{aligned} B_{11} &= 4 \frac{1 + \sigma_R^2}{1 - \sigma_R^2} \|I_N - \mathbf{1}\pi_{\mathbf{R}}^T\|_R^2 \|\pi_{\mathbf{C}}\|_R^2, \\ B_{12} &= Nm \|\mathbf{R} - \mathbf{1}\pi_{\mathbf{R}}^T\|_R^2, \quad B_{13} = Nm \|(I - \mathbf{1}\pi_{\mathbf{R}}^T)\mathbf{C}\|_R^2, \\ B_{21} &= 2 \|\pi_{\mathbf{C}}\|_C^2 \mathbf{a}_4, \quad B_{22} = \frac{1}{3} \|\mathbf{R}\|_C^2 N m \mathbf{a}_5, \\ B_{23,k} &= Nm \left[\|2(\mathbf{C}_\gamma - I)\mathbf{C}\|_C^2 + 4L\alpha_{k+1} \right. \\ &\quad \left. + L^2 \|\mathbf{C}\|_C \max\{\alpha_{k+1}^2, \alpha_k^2\} \right] \mathbf{a}_6, \end{aligned}$$

with the constants in the following:

$$\begin{aligned} \mathbf{a}_1 &= 4 \frac{1 + \sigma_R^2}{1 - \sigma_R^2} \|I_N - \mathbf{1}\pi_{\mathbf{R}}^T\|_R^2 \|\pi_{\mathbf{C}}\|_R^2 L^2 N, \\ \mathbf{a}_2 &= 2 \frac{1 + \sigma_R^2}{1 - \sigma_R^2} \|I_N - \mathbf{1}\pi_{\mathbf{R}}^T\|_R^2 \delta_{R,C}^2, \\ \mathbf{a}_3 &= 3 \frac{1 + \sigma_C^2}{1 - \sigma_C^2} \|I - \pi_{\mathbf{C}}\mathbf{1}^T\|_C^2 \|\mathbf{R}_\phi - I\|_C^2 L^2 \delta_{C,R}^2, \\ \mathbf{a}_4 &= 6 \frac{1 + \sigma_C^2}{1 - \sigma_C^2} \|I - \pi_{\mathbf{C}}\mathbf{1}^T\|_C^2 L^2 \delta_{C,R}^2, \\ \mathbf{a}_5 &= 3 \frac{1 + \sigma_C^2}{1 - \sigma_C^2} \|I - \pi_{\mathbf{C}}\mathbf{1}^T\|_C^2 L^2, \\ \mathbf{a}_6 &= \frac{1 + \sigma_C^2}{1 - \sigma_C^2} \|I - \pi_{\mathbf{C}}\mathbf{1}^T\|_C^2. \end{aligned}$$

Proof. The proof is provided in Appendix A.2 (Huo et al., 2024).

Note that when the step size $\alpha_k \rightarrow 0$, the matrix \mathbf{P}_k tends to become upper-triangular. Its eigenvalues approach $q_R = \frac{1 + \sigma_R^2}{2}$ and $q_C = \frac{1 + \sigma_C^2}{2}$, where $\sigma_R < 1$ and $\sigma_C < 1$ are defined in Lemma 2. In the following, we establish conditions regarding step sizes and variances of Laplacian noise for the convergence of (15) with non-convex $f_i, \forall i \in \mathcal{N}$.

Theorem 1. Under Assumptions 2–4 and the L -Lipschitz smoothness of $f_i, \forall i \in \mathcal{N}$, when $\sum_{k=0}^{\infty} \alpha_k < \infty, \sum_{k=0}^{\infty} \theta_{\xi,k}^2 < \infty, \sum_{k=0}^{\infty} \theta_{\zeta,k}^2 < \infty, \sum_{k=0}^{\infty} \frac{\theta_{\xi,k}^2}{\alpha_k} < \infty, \sum_{k=0}^{\infty} \frac{\theta_{\zeta,k}^2}{\alpha_k} < \infty$, and there exists λ satisfying $q_R < \lambda < 1$ and $q_C < \lambda < 1$ and $k_0 > 0$ such that $\frac{\alpha_k}{\alpha_0} \geq \beta \lambda^{k-k_0}$, we have that

- (i) $\lim_{k \rightarrow \infty} \mathbb{E}[\|\mathbf{x}_k - \mathbf{1}_N \bar{\mathbf{x}}_k^T\|] = 0$.
- (ii) $\mathbb{E}[f(\bar{\mathbf{x}}_k)] - f^*$ converges to a finite value almost surely.

Proof. We first bound $\|\bar{\mathbf{v}}_k\|^2$. According to the definition of $\bar{\mathbf{v}}_k$, we have

$$\begin{aligned} \bar{\mathbf{v}}_k &= (\mathbf{v}_k - \pi_{\mathbf{C}} \hat{\mathbf{v}}_k^T + \pi_{\mathbf{C}} \hat{\mathbf{v}}_k^T)^T \pi_{\mathbf{R}} \\ &= (\mathbf{v}_k - \pi_{\mathbf{C}} \hat{\mathbf{v}}_k^T)^T \pi_{\mathbf{R}} + \alpha_k (G(\mathbf{x}_k) - G(\mathbf{1}\bar{\mathbf{x}}_k^T))^T \mathbf{1}\pi_{\mathbf{C}}^T \pi_{\mathbf{R}} \\ &\quad + \alpha_k \nabla f(\bar{\mathbf{x}}_k) \pi_{\mathbf{C}}^T \pi_{\mathbf{R}}, \end{aligned}$$

and hence,

$$\begin{aligned} \|\bar{\mathbf{v}}_k\|^2 &\leq 3 \|\pi_{\mathbf{R}}\|^2 \delta_{2,C}^2 \|\mathbf{v}_k - \pi_{\mathbf{C}} \hat{\mathbf{v}}_k^T\|_C^2 \\ &\quad + 3L^2 N (\pi_{\mathbf{C}}^T \pi_{\mathbf{R}})^2 \delta_{2,R}^2 \alpha_k^2 \|\mathbf{x}_k - \mathbf{1}\bar{\mathbf{x}}_k^T\|_R^2 \\ &\quad + 3(\pi_{\mathbf{C}}^T \pi_{\mathbf{R}})^2 \alpha_k^2 \|\nabla f(\bar{\mathbf{x}}_k)\|^2. \end{aligned}$$

For $-\nabla f(\bar{\mathbf{x}}_k)^T \bar{\mathbf{v}}_k$, we have

$$\begin{aligned} &-\nabla f(\bar{\mathbf{x}}_k)^T \bar{\mathbf{v}}_k \\ &\leq -\pi_{\mathbf{C}}^T \pi_{\mathbf{R}} \alpha_k \|\nabla f(\bar{\mathbf{x}}_k)\|^2 + \|\pi_{\mathbf{R}}\| \|\nabla f(\bar{\mathbf{x}}_k)\| \|\mathbf{v}_k - \pi_{\mathbf{C}} \hat{\mathbf{v}}_k^T\| \end{aligned}$$

$$+ L\sqrt{N} \pi_{\mathbf{C}}^T \pi_{\mathbf{R}} \alpha_k \|\nabla f(\bar{\mathbf{x}}_k)\| \|\mathbf{x}_k - \mathbf{1}\bar{\mathbf{x}}_k^T\|.$$

Since f_i is L -Lipschitz smooth, we have $f(\bar{\mathbf{x}}_{k+1}) \leq f(\bar{\mathbf{x}}_k) + (\bar{\mathbf{x}}_{k+1} - \bar{\mathbf{x}}_k)^T \nabla f(\bar{\mathbf{x}}_k) + \frac{L}{2} \|\bar{\mathbf{x}}_{k+1} - \bar{\mathbf{x}}_k\|^2$. According to $\bar{\mathbf{x}}_{k+1} - \bar{\mathbf{x}}_k = (\mathbf{x}_{k+1} - \mathbf{x}_k)^T \pi_{\mathbf{R}} = -\mathbf{v}_k^T \pi_{\mathbf{R}} - \phi \zeta_k^T \pi_{\mathbf{R}} - \gamma (\mathbf{C} \xi_k)^T \pi_{\mathbf{R}}$, we have

$$\begin{aligned} &\mathbb{E}[f(\bar{\mathbf{x}}_{k+1})] \\ &\leq \mathbb{E}[f(\bar{\mathbf{x}}_k)] - \mathbb{E}[\nabla f(\bar{\mathbf{x}}_k)^T (\bar{\mathbf{v}}_k + \phi \zeta_k^T \pi_{\mathbf{R}} + \gamma (\mathbf{C} \xi_k)^T \pi_{\mathbf{R}})] \\ &\leq \mathbb{E}[f(\bar{\mathbf{x}}_k)] + \frac{L}{2} \|\pi_{\mathbf{R}}\|^2 \phi^2 N m \theta_{\xi,k}^2 \\ &\quad + \frac{L}{2} \|\mathbf{C}^T \pi_{\mathbf{C}}\|^2 \gamma^2 N m \theta_{\xi,k}^2 \\ &\quad + \frac{\delta_{2,R}^2}{2} (\pi_{\mathbf{C}}^T \pi_{\mathbf{R}})^2 (3L^3 N \alpha_k^2 + 1) \|\mathbf{x}_k - \mathbf{1}\bar{\mathbf{x}}_k^T\|_R^2 \\ &\quad + \frac{\delta_{2,C}^2}{2\alpha_k \pi_{\mathbf{C}}^T \pi_{\mathbf{R}}} \|\pi_{\mathbf{R}}\|^2 (3L + 1) \|\mathbf{v}_k - \pi_{\mathbf{C}} \hat{\mathbf{v}}_k^T\|_C^2 \\ &\quad - \left(\frac{1}{2} \pi_{\mathbf{C}}^T \pi_{\mathbf{R}} \alpha_k \right. \\ &\quad \left. - \frac{3L(\pi_{\mathbf{C}}^T \pi_{\mathbf{R}})^2 + L^2 N}{2} \alpha_k^2 \right) \mathbb{E}[\|\nabla f(\bar{\mathbf{x}}_k)\|^2]. \end{aligned} \quad (19)$$

Since $\sum_{k=0}^{\infty} \alpha_k < \infty$, there exists K such that $\forall k \geq K$, one has $\alpha_k < \frac{\pi_{\mathbf{C}}^T \pi_{\mathbf{R}}}{3L(\pi_{\mathbf{C}}^T \pi_{\mathbf{R}})^2 + 1 + L^2 N}$. For $k \leq K$, there always exists a bound for $\mathbb{E}[\|\nabla f(\bar{\mathbf{x}}_k)\|^2]$, $\mathbb{E}[\|\mathbf{x}_k - \mathbf{1}\bar{\mathbf{x}}_k^T\|_R^2]$, and $\mathbb{E}[\|\mathbf{v}_k - \pi_{\mathbf{C}} \hat{\mathbf{v}}_k^T\|_C^2]$. Then we only need to prove the boundedness of these three values for $k > K$. Specifically, denoting $g_k = \mathbb{E}[\|\nabla f(\bar{\mathbf{x}}_k)\|^2]$, $X_k = \mathbb{E}[\|\mathbf{x}_k - \mathbf{1}\bar{\mathbf{x}}_k^T\|_R^2]$, and $V_k = \mathbb{E}[\|\mathbf{v}_k - \pi_{\mathbf{C}} \hat{\mathbf{v}}_k^T\|_C^2]$ when $k \geq K$. We will prove the following:

$$g_k \leq D_g, \quad X_k \leq D_X, \quad V_k \leq D_V, \quad \forall k > K, \quad (20)$$

where $D_g, D_X, D_V > 0$ are some constants. We prove (20) by induction. Assume that (20) holds form certain $k > 0$, then we need to prove that

$$\begin{aligned} X_{k+1} &\leq q_R X_k + \mathbf{a}_1 \alpha_k^2 D_X + \mathbf{a}_2 V_k + B_{11} \alpha_k^2 D_g \\ &\quad + B_{12} \phi^2 \theta_{\zeta,k}^2 + B_{13} \gamma^2 \theta_{\xi,k}^2, \end{aligned} \quad (21a)$$

$$V_{k+1} \leq q_C V_k + r_{1,k}, \quad (21b)$$

with $r_{1,k} = \mathbf{a}_5 \max\{\alpha_{k+1}^2, \alpha_k^2\} D_V + (\mathbf{a}_3 + \mathbf{a}_4 \alpha_k^2) \max\{\alpha_{k+1}^2, \alpha_k^2\} D_X + B_{21} \alpha_k^2 D_g + B_{22} \phi^2 \theta_{\zeta,k}^2 + B_{23} \gamma^2 \theta_{\xi,k}^2$, and then (21b) suffices to show $V_{k+1} \leq q_C^{k+1} V_0 + \sum_{l=0}^k q_C^{k-l} r_{1,l} \leq D_V$. Since $\sum_{l=0}^k r_{1,l} \leq \sum_{l=0}^{\infty} r_{1,l}$ and

$$\begin{aligned} \sum_{l=0}^{\infty} r_{1,l} &\leq (\mathbf{a}_5 D_V + B_{21} D_g + \mathbf{a}_3 D_X) \sum_{l=0}^{\infty} \alpha_l^2 \\ &\quad + B_{22} \phi^2 \sum_{l=0}^{\infty} \theta_{\zeta,k}^2 + B_{23} \gamma^2 \sum_{l=0}^{\infty} \theta_{\xi,k}^2 \\ &\quad + o\left(\sum_{l=0}^{\infty} \alpha_l^2 + \sum_{l=0}^{\infty} \theta_{\xi,k}^2 + \sum_{l=0}^{\infty} \theta_{\zeta,k}^2\right) \\ &= D'_V < \infty, \end{aligned}$$

by defining $D_V = V_0 + D'_V$, V_0 always satisfies $V_0 \leq D_V$. By induction and $q_C < 1$, we have $V_{k+1} < V_0 + \sum_{l=0}^k r_{1,l} \leq D_V$.

Since $\frac{\alpha_k}{\alpha_0} \leq \beta \lambda^{k-k_0}$, one further has $V_{k+1} \leq q_C^{k+1} V_0 + \sum_{l=0}^k q_C^{k-l} r_{1,l} \leq \left(\frac{q_C}{\lambda}\right)^{k+1} \alpha_{k+1} \frac{V_0}{\beta \alpha_0} + \frac{\alpha_{k+1}}{\beta \lambda} \sum_{l=0}^k \left(\frac{q_C}{\lambda}\right)^{k-l} \frac{r_{1,l}}{\alpha_l}$, which further yields that

$$\sum_{k=0}^{\infty} \frac{V_k}{\alpha_k} \leq \frac{\lambda}{\lambda - q_C} \frac{V_0}{\beta \alpha_0} + \frac{1}{\beta(\lambda - q_C)} \sum_{k=0}^{\infty} \frac{r_{1,k}}{\alpha_k}$$

$$\begin{aligned}
 &\leq \frac{\lambda}{\lambda - q_C} \frac{V_0}{\beta \alpha_0} + \frac{2\mathbf{a}_5 D_V}{\beta(\lambda - q_C)} \sum_{l=0}^{\infty} \alpha_l \\
 &\quad + \frac{B_{22} \phi^2}{\beta(\lambda - q_C)} \sum_{l=0}^{\infty} \frac{\theta_{\xi,k}^2}{\alpha_k^2} \\
 &\quad + \frac{Nm \|2(\mathbf{C}_\gamma - I)\mathbf{C}\|_C^2 \gamma^2}{\beta(\lambda - q_C)} \sum_{l=0}^{\infty} \frac{\theta_{\xi,k}^2}{\alpha_k^2} \\
 &\quad + o\left(\sum_{l=0}^{\infty} \alpha_l + \sum_{l=0}^{\infty} \frac{\theta_{\xi,k}^2}{\alpha_k}\right) < \infty. \tag{22}
 \end{aligned}$$

Therefore, we can infer $\sum_{k=0}^{\infty} V_k < \infty$.

Define $r_{2,k} = \mathbf{a}_1 \alpha_k^2 D_X + \mathbf{a}_2 V_k + B_{11} \alpha_k^2 D_g + B_{12} \phi^2 \theta_{\xi,k}^2 + B_{13} \gamma^2 \theta_{\xi,k}^2$, and then (21a) suffices to show $X_{k+1} \leq q_R^{k+1} X_0 + \sum_{l=0}^k q_R^{k-l} r_{2,l} \leq D_X$. We have

$$\begin{aligned}
 \sum_{k=0}^{\infty} r_{2,k} &\leq \sum_{k=0}^{\infty} (\mathbf{a}_1 D_X + B_{11} D_g) \alpha_k^2 + \mathbf{a}_2 \sum_{k=0}^{\infty} V_k \\
 &\quad + B_{12} \phi^2 \sum_{k=0}^{\infty} \theta_{\xi,k}^2 + B_{13} \gamma^2 \sum_{k=0}^{\infty} \theta_{\xi,k}^2 \\
 &= D'_X < \infty.
 \end{aligned}$$

By letting $D_X = X_0 + D'_X$, we have $X_{k+1} < X_0 + \sum_{l=0}^k r_{2,l} \leq D_X$.

Similar to (22), one has

$$\begin{aligned}
 \sum_{k=0}^{\infty} \frac{X_k}{\alpha_k} &\leq \frac{\lambda}{\lambda - q_R} \frac{X_0}{\beta \alpha_0} + \frac{1}{\beta(\lambda - q_R)} \sum_{k=0}^{\infty} \frac{r_{2,k}}{\alpha_k} \\
 &\leq \frac{\lambda}{\lambda - q_R} \frac{X_0}{\beta \alpha_0} + \frac{\mathbf{a}_1 D_X + B_{11} D_g}{\lambda - q_R} \sum_{k=0}^{\infty} \alpha_k + \frac{\mathbf{a}_2}{\lambda - q_R} \sum_{k=0}^{\infty} \frac{V_k}{\alpha_k} \\
 &\quad + \frac{B_{12} \phi^2}{\lambda - q_R} \sum_{k=0}^{\infty} \frac{\theta_{\xi,k}^2}{\alpha_k} + \frac{B_{13} \gamma^2}{\lambda - q_R} \sum_{k=0}^{\infty} \frac{\theta_{\xi,k}^2}{\alpha_k} < \infty. \tag{23}
 \end{aligned}$$

Hence, we have $\sum_{k=0}^{\infty} X_k < \infty$, and $\lim_{k \rightarrow \infty} X_k = 0$ since $X_k \geq 0$. According to (19), we have

$$\begin{aligned}
 \mathbb{E}[f(\bar{\mathbf{x}}_{k+1})] - f^* &\leq \mathbb{E}[f(\bar{\mathbf{x}}_k)] - f^* \\
 &\quad - \left(\alpha_k \pi_C^T \pi_R - \frac{3L(\pi_C^T \pi_R)^2 + 1 + L^2 N}{2} \alpha_k^2 \right) \mathbb{E}[\|\nabla f(\bar{\mathbf{x}}_k)\|^2] \\
 &\quad + r_{3,k}, \tag{24}
 \end{aligned}$$

where $r_{3,k} = \frac{1}{2} \|\pi_R\|^2 \phi^2 Nm \theta_{\xi,k}^2 + \frac{1}{2} \|\mathbf{C}^T \pi_C\|^2 \gamma^2 Nm \theta_{\xi,k}^2 + \frac{\delta_{2,R}^2}{2} (\pi_C^T \pi_R)^2 (3L^3 N \alpha_k^2 + 1) X_k + \frac{\delta_{2,C}^2}{2} \|\pi_R\|^2 (3L + 1) \frac{V_k}{\alpha_k}$. Since $\sum_{k=0}^{\infty} \theta_{\xi,k}^2 < \infty$, $\sum_{k=0}^{\infty} \theta_{\xi,k}^2 < \infty$, $\sum_{k=0}^{\infty} \frac{V_k}{\alpha_k} < \infty$, and $\sum_{k=0}^{\infty} X_k < \infty$, we have $\sum_{k=0}^{\infty} r_{3,k} < \infty$. Furthermore, $\alpha_k < \frac{2\pi_C^T \pi_R}{3L(\pi_C^T \pi_R)^2 + 1 + L^2 N}$ yields the non-negativeness of the third term in (24).

Based on Lemma 5 in Appendix A.1 (Huo et al., 2024), we conclude that $\mathbb{E}[f(\bar{\mathbf{x}}_k)] - f^*$ converges a.s. to a finite value and

$$\sum_{k=k'}^{\infty} \left(\alpha_k \pi_C^T \pi_R - \frac{3L(\pi_C^T \pi_R)^2 + 1 + L^2 N}{2} \alpha_k^2 \right) \mathbb{E}[\|\nabla f(\bar{\mathbf{x}}_k)\|^2] < \infty.$$

Therefore, we conclude that $\sup_k \mathbb{E}[\|\nabla f(\bar{\mathbf{x}}_k)\|^2]$ exists, and thus we can define $D_g = \sup_k \mathbb{E}[\|\nabla f(\bar{\mathbf{x}}_k)\|^2]$. Here, we complete the proof of (20) and additionally prove that $\lim_{k \rightarrow \infty} \mathbb{E}[\|\mathbf{x}_k - \mathbf{1} \bar{\mathbf{x}}_k^T\|_R^2] = 0$, and $\mathbb{E}[f(\bar{\mathbf{x}}_k)] - f^*$ converges a.s. to a finite value.

Remark 3. As shown in (22), (23), and (24), the convergence rate of $\mathbb{E}[f(\bar{\mathbf{x}}_k)] - f^*$ depends on the decay rates of $\sum_{k=0}^{\infty} \alpha_k$,

$\sum_{k=0}^{\infty} \theta_{\xi,k}^2$, $\sum_{k=0}^{\infty} \theta_{\zeta,k}^2$, $\sum_{k=0}^{\infty} \frac{\theta_{\xi,k}^2}{\alpha_k}$, and $\sum_{k=0}^{\infty} \frac{\theta_{\zeta,k}^2}{\alpha_k}$ under $\sup_k \alpha_k < 1$. For example, choosing parameters $(\alpha_k, \theta_{\xi,k}, \theta_{\zeta,k})$ as $O(1/k^{1+p})$ for any $p > 0$ yields an $O(1/k^p)$ convergence rate for DP-DGT.

Remark 4. Theorem 1 depends only on the smoothness of f_i , not its convexity, extending the analysis in Chen et al. (2023), Ding et al. (2021b), Huang et al. (2024), Pu (2020), Wang and Nedić (2023). This allows the method to apply to distributed non-convex optimization with information-sharing noise. Additionally, Huang et al. (2024, Theorem 2) shows that the gradient tracking algorithm with robust push-pull in (15) cannot achieve ϵ -differential privacy if step sizes are not summable under Laplacian noise. Therefore, the convergence analysis with constant step sizes (Zhang et al., 2020) is inapplicable. We provide a rigorous convergence analysis for summable step size sequences under Laplace noise, subject to specific conditions.

4.2. Convergence of primal variables

Since f_i in (5) is convex, Theorem 1 ensures that under DP-DGT, the dual variables converge to a neighborhood of the optimal solution of (5). Thus, we can now establish the convergence of DP-DGT for the distributed resource allocation problem (1).

Theorem 2. Under Assumptions 1–4, if $\sum_{k=0}^{\infty} \theta_{\xi,k}^2 < \infty$, $\sum_{k=0}^{\infty} \theta_{\zeta,k}^2 < \infty$, $\sum_{k=0}^{\infty} \alpha_k < \infty$, $\sum_{k=0}^{\infty} \frac{\theta_{\xi,k}^2}{\alpha_k} < \infty$, $\sum_{k=0}^{\infty} \frac{\theta_{\zeta,k}^2}{\alpha_k} < \infty$, and there exists λ satisfying $q_R < \lambda < 1$ and $q_C < \lambda < 1$ and $k_0 > 0$ such that $\frac{\alpha_k}{\alpha_{k_0}} \geq \beta \lambda^{k-k_0}$, then the sequence $\{\mathbf{w}_k\}$ in Algorithm 1 converges to a neighborhood of \mathbf{w}^* , where $\mathbf{w}_k = [w_{1,k}, \dots, w_{N,k}]^T \in \mathbb{R}^{N \times m}$ and $\mathbf{w}^* = [w_1^*, \dots, w_N^*]^T \in \mathbb{R}^{N \times m}$.

Proof. The proof is provided in Appendix A.3 (Huo et al., 2024).

Our proposed algorithm exhibits constraint violations due to inherent errors in the dual variable solution, as shown in Theorem 1. This issue is common in works using Laplacian noise, as DP noise can affect convergence accuracy and constraint satisfaction. While some methods achieve accurate convergence (Munoz et al., 2021; Wang & Nedić, 2024), they are limited to specific problem types or privacy aspects. In conclusion, constrained distributed optimization faces challenges in balancing DP and constraint satisfaction, and we aim to develop advanced algorithms to address these violations in the future.

5. Differential privacy analysis

In our analysis, we consider the worst-case scenario where the adversary can observe all communication in the network and has access to the initial value of the algorithm. Thus, we denote the attacker's observation sequence as $\{\mathcal{O}_k\}_{k \geq 0}$, and the observation at time k is $\mathcal{O}_k = \{s_{i,k} + \xi_{i,k}, \tilde{w}_{i,k} + \zeta_{i,k} \mid \forall i \in \mathcal{N}\}$.

Theorem 3. Under Assumptions 1–4, if $\sum_{k=0}^{\infty} \alpha_k < \infty$, $D_{\alpha,\xi} := \sum_{k=0}^{\infty} \frac{\alpha_k}{\theta_{\xi,k}} < \infty$, $D_{\alpha,\zeta} := \sum_{k=0}^{\infty} \frac{\alpha_k}{\theta_{\zeta,k}} < \infty$, and $\pi_C^T \pi_R < \frac{1}{2}$, then Algorithm 1 achieves ϵ -differential privacy for any two δ -adjacent distributed resource allocation problems, with the cumulative privacy budget

$$\epsilon = \frac{\delta + D_\eta}{\mu \gamma \phi} (D_{\alpha,\xi} + \phi D_{\alpha,\zeta}), \tag{25}$$

where

$$D_\eta = \inf_{K \geq k_0} \max \left\{ \frac{\max_{0 \leq l < K} \{\alpha_l \delta + \alpha_l \eta\} + \bar{\alpha} \delta}{\gamma \phi \mu - \bar{\alpha}}, \max_{0 \leq l < K} \eta_l \right\}, \tag{26}$$

with

$$\bar{\alpha} = \sup_k \alpha_k \text{ and } k_0 = \min_k \{\alpha_t < \mu\gamma\phi\}.$$

Proof. We consider the implementation of the proposed algorithm for both resource allocation problem \mathcal{P} and \mathcal{P}' . Since it is assumed that the attacker knows all auxiliary information, including the initial states, local demands, and the network topology, we have $\mathbf{s}_0 = \mathbf{s}'_0$, $\tilde{\mathbf{w}}_0 = \tilde{\mathbf{w}}'_0$, and $\mathbf{w}_0 = \mathbf{w}'_0$. From Algorithm 1, it can be seen that given initial state $\{\mathbf{s}_{i,0}, \tilde{\mathbf{w}}_0, \mathbf{w}_0\}$, the communication graphs $\{\mathbf{R}, \mathbf{C}\}$ and the function set and the observation sequence $\mathcal{O} = \{\mathcal{O}_k\}$ is uniquely determined by the noise sequences $\{\xi_k\}$ and $\{\zeta_k\}$. Thus, it is equivalent to prove that

$$\mathbb{P}[\mathcal{R}^{-1}(\mathcal{P}, \mathcal{O}, \mathbf{s}_0, \tilde{\mathbf{w}}_0, \mathbf{w}_0)] \leq e^\epsilon \mathbb{P}[\mathcal{R}^{-1}(\mathcal{P}', \mathcal{O}, \mathbf{s}_0, \tilde{\mathbf{w}}_0, \mathbf{w}_0)].$$

The attacker can eavesdrop on the transmitted messages, and therefore, there is $\mathbf{s}_k + \xi_k = \mathbf{s}'_k + \xi'_k$, and $\tilde{\mathbf{w}}_k + \zeta_k = \tilde{\mathbf{w}}'_k + \zeta'_k$.

For any $i \neq i_0$, since $s_{i,0} = s'_{i,0}$, $w_{i,0} = w'_{i,0}$, and $s_{j,0} + \xi_{j,0} = s'_{j,0} + \xi'_{j,0}$ for any $j \in \mathcal{N}$, we obtain $s_{i,1} = s'_{i,1}$ based on (12a). Due to $\tilde{w}_{i,0} = \tilde{w}'_{i,0}$ and $\tilde{w}_{j,0} + \zeta_{j,0} = \tilde{w}'_{j,0} + \zeta'_{j,0}$ for any $j \in \mathcal{N}$, it then follows that $\tilde{w}_{i,1} = \tilde{w}'_{i,1}$ according to (12c). Also, $w_{i,1} = w'_{i,1}$ can be inferred from (12b) due to $F_i = F'_i$. Based on the above analysis, we can finally obtain

$$\xi_{i,k} = \xi'_{i,k}, \zeta_{i,k} = \zeta'_{i,k}, \forall k \geq 0, \forall i \neq i_0. \quad (27)$$

For agent i_0 , the noise should satisfy

$$\Delta \xi_{i_0,k} = -\Delta s_{i_0,k}, \Delta \zeta_{i_0,k} = -\Delta \tilde{w}_{i_0,k}, \forall k > 0, \quad (28)$$

where $\Delta \xi_{i_0,k} = \xi_{i_0,k} - \xi'_{i_0,k}$, $\Delta s_{i_0,k} = s_{i_0,k} - s'_{i_0,k}$, $\Delta \zeta_{i_0,k} = \zeta_{i_0,k} - \zeta'_{i_0,k}$, and $\Delta \tilde{w}_{i_0,k} = \tilde{w}_{i_0,k} - \tilde{w}'_{i_0,k}$. We have

$$\begin{aligned} \Delta s_{i_0,k+1} &= (1-\gamma)\Delta s_{i_0,k} - \alpha_k \Delta w_{i_0,k}, \\ \Delta \tilde{w}_{i_0,k+1} &= (1-\phi)\Delta \tilde{w}_{i_0,k} + (\Delta s_{i_0,k+1} - \Delta s_{i_0,k}), \end{aligned} \quad (29)$$

where $\Delta w_{i_0,k} = w_{i_0,k} - w'_{i_0,k}$. Based on the primal variable update (12b), the following relationship holds:

$$\begin{aligned} \|\tilde{w}_{i_0,k} - \tilde{w}'_{i_0,k}\|_1 &= \|\nabla F_{i_0}(w_{i_0,k}) - \nabla F'_{i_0}(w'_{i_0,k})\|_1 \\ &= \|\nabla F_{i_0}(w_{i_0,k}) - \nabla F_{i_0}(w'_{i_0,k}) \\ &\quad + \nabla F_{i_0}(w'_{i_0,k}) - \nabla F'_{i_0}(w'_{i_0,k})\|_1 \\ &\geq \|\nabla F_{i_0}(w_{i_0,k}) - \nabla F_{i_0}(w'_{i_0,k})\|_1 - \delta \\ &\geq \mu \|\Delta w_{i_0,k}\|_1 - \delta, \end{aligned}$$

where the first inequality is from Definition 2. Therefore, we have

$$\|\Delta w_{i_0,k}\|_1 \leq \frac{\|\Delta \tilde{w}_{i_0,k}\|_1}{\mu} + \frac{\delta}{\mu}. \quad (30)$$

Then, taking the ℓ_1 -norm of both side of (29) yields

$$\begin{aligned} \|\Delta s_{i_0,k+1}\|_1 &\leq (1-\gamma)\|\Delta s_{i_0,k}\|_1 + \frac{\alpha_k}{\mu} \|\Delta \tilde{w}_{i_0,k}\|_1 + \frac{\alpha_k \delta}{\mu}, \\ \|\Delta \tilde{w}_{i_0,k+1}\|_1 &\leq \left(1-\phi + \frac{\alpha_k}{\mu}\right) \|\Delta \tilde{w}_{i_0,k}\|_1 \\ &\quad + (2-\gamma)\|\Delta s_{i_0,k}\|_1 + \frac{\alpha_k \delta}{\mu}. \end{aligned}$$

Consider a discrete-time dynamical system,

$$\begin{bmatrix} \varphi_{k+1} \\ \eta_{k+1} \end{bmatrix} = \begin{bmatrix} 1-\gamma & \frac{\alpha_k}{\mu} \\ 2-\gamma & 1-\phi + \frac{\alpha_k}{\mu} \end{bmatrix} \begin{bmatrix} \varphi_k \\ \eta_k \end{bmatrix} + \frac{\alpha_k \delta}{\mu} \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad (31)$$

with $\varphi_0 = 0$ and $\eta_0 = 0$. Since $\|\Delta s_{i_0,0}\|_1 \leq \varphi_0$ and $\|\Delta \tilde{w}_{i_0,0}\|_1 \leq \eta_0$, we infer that $\|\Delta s_{i_0,k}\|_1 \leq \varphi_k$ and $\|\Delta \tilde{w}_{i_0,k}\|_1 \leq \eta_k$, $\forall k \geq 0$, by induction. Moreover, we have $\begin{bmatrix} \varphi_{k+1} \\ \eta_{k+1} \end{bmatrix} = \sum_{l=0}^k \begin{bmatrix} \alpha_l \delta \\ \alpha_l \eta_l \end{bmatrix} P^{k-l}$

$$\begin{aligned} &\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \text{ with } P^0 = I \text{ and } P^k = \\ &\begin{bmatrix} (1-\gamma)^k & 0 \\ \sum_{l=0}^{k-1} (2-\gamma)(1-\phi)^{k-1-l}(1-\gamma)^l & (1-\phi)^k \end{bmatrix} \text{ for } k > 0. \text{ Hence,} \\ \varphi_{k+1} &= \sum_{l=0}^k (1-\gamma)^{k-l} \left(\frac{\alpha_l \delta}{\mu} + \frac{\alpha_l \eta_l}{\mu} \right), \quad (32) \\ \eta_{k+1} &= \sum_{l=0}^k \left[\sum_{j=0}^{k-l-1} (2-\gamma)(1-\phi)^{k-1-j} \left(\frac{1-\gamma}{1-\phi} \right)^j \right. \\ &\quad \left. + (1-\phi)^{k-l} \right] \left(\frac{\alpha_l \delta}{\mu} + \frac{\alpha_l \eta_l}{\mu} \right) \\ &= \sum_{l=0}^k \left[(2-\gamma) \frac{(1-\phi)^{k-l} - (1-\gamma)^{k-l}}{\gamma-\phi} \right. \\ &\quad \left. + (1-\phi)^{k-l} \right] \left(\frac{\alpha_l \delta}{\mu} + \frac{\alpha_l \eta_l}{\mu} \right) \\ &= \frac{1}{\gamma-\phi} \sum_{l=0}^k \left[(2-\phi)(1-\phi)^{k-l} \right. \\ &\quad \left. - (2-\gamma)(1-\gamma)^{k-l} \right] \left(\frac{\alpha_l \delta}{\mu} + \frac{\alpha_l \eta_l}{\mu} \right). \quad (33) \end{aligned}$$

We first prove that η_k is bounded, i.e., $\eta_k \leq D_\eta$. We separate the sequence $\{\eta_k\}_{k \geq 0}$ into two part. One is for $k < K$, and the other one is for $k \geq K$. For the first part, there always exists a bound for $\{\eta_k\}_{0 \leq k < K}$ since it only has a finite number of elements. Therefore, we only need to prove the boundness of $\{\eta_k\}_{k \geq K}$. We prove it by induction. Suppose that there exists an $k \geq K$ such that $\eta_l \leq D'_\eta$ for all $K \leq l \leq k$, and it is sufficient to prove that $\eta_{k+1} \leq D'_\eta$ by induction.

We write (33) as

$$\begin{aligned} \eta_{k+1} &= \frac{1}{\gamma-\phi} \sum_{l=0}^{K-1} \left[(2-\phi)(1-\phi)^{k-l} \right. \\ &\quad \left. - (2-\gamma)(1-\gamma)^{k-l} \right] \left(\frac{\alpha_l \delta}{\mu} + \frac{\alpha_l \eta_l}{\mu} \right) \\ &\quad + \frac{1}{\gamma-\phi} \sum_{l=K}^k \left[(2-\phi)(1-\phi)^{k-l} \right. \\ &\quad \left. - (2-\gamma)(1-\gamma)^{k-l} \right] \left(\frac{\alpha_l \delta}{\mu} + \frac{\alpha_l \eta_l}{\mu} \right). \quad (34) \end{aligned}$$

For the first term of (34), there is

$$\begin{aligned} &\sum_{l=0}^{K-1} \left[(2-\phi)(1-\phi)^{k-l} - (2-\gamma)(1-\gamma)^{k-l} \right] \left(\frac{\alpha_l \delta}{\mu} + \frac{\alpha_l \eta_l}{\mu} \right) \\ &\leq \frac{2(\gamma-\phi)}{\gamma\phi\mu} \max_{0 \leq l < K} \{\alpha_l \delta + \alpha_l \eta_l\}. \end{aligned}$$

Thus, we define $D''_\eta := \frac{2 \max_{0 \leq l < K} \{\alpha_l \delta + \alpha_l \eta_l\}}{\gamma\phi\mu}$. Since $\sum_{k=0}^{\infty} \alpha_k < \infty$, there exists a finite $\bar{\alpha}$ such that $\bar{\alpha} = \sup_k \alpha_k$. Define $D'_\eta = \frac{\mu\gamma\phi D''_\eta + \bar{\alpha}\delta}{\mu\gamma\phi - \bar{\alpha}}$, then we have the following relationship based on (34)

$$\eta_{k+1} \leq D''_\eta + \frac{\bar{\alpha}(\delta + D'_\eta)}{\mu\gamma\phi} = D'_\eta.$$

Therefore, we can conclude that $\eta_k \leq D_\eta$ by letting $D_\eta = \inf_K \max \{D'_\eta, \max_{0 \leq l < K} \eta_l\}$, $\forall k \geq 0$.

Since $\eta_k \leq D_\eta$, we have the following result:

$$\sum_{k=0}^T \frac{\|\Delta s_{i_0,k}\|_1}{\theta_{\xi,k}} \leq \sum_{k=0}^T \frac{\varphi_k}{\theta_{\xi,k}}$$

$$\begin{aligned}
 &= \sum_{k=1}^T \frac{1}{\theta_{\xi,k}} \sum_{l=0}^{k-1} (1-\gamma)^{k-1-l} \left(\frac{\alpha_l \delta}{\mu} + \frac{\alpha_l \eta_l}{\mu} \right) \\
 &\leq \sum_{k=0}^{T-1} \frac{\alpha_k \delta + \alpha_k \eta_k}{\mu \theta_{\xi,k+1}} \sum_{l=0}^{T-1-k} (1-\gamma)^l \\
 &\leq \frac{\delta + D_\eta}{\mu \gamma} \sum_{k=0}^{T-1} \frac{\alpha_k}{\theta_{\xi,k+1}} \\
 &\leq \frac{\delta + D_\eta}{\mu \gamma} D_{\alpha,\xi} < \infty, \quad \forall T \geq 0.
 \end{aligned} \tag{35}$$

Similarly, we have

$$\begin{aligned}
 &\sum_{k=0}^T \frac{\|\Delta \tilde{w}_{i_0,k}\|}{\theta_{\zeta,k}} \leq \sum_{k=0}^T \frac{\eta_k}{\theta_{\zeta,k}} \\
 &\leq \sum_{k=1}^T \frac{\alpha_k \delta + \alpha_k \eta_k}{(\gamma - \phi) \mu \theta_{\zeta,k}} \sum_{l=0}^{T-1-k} [(2-\phi)(1-\phi)^{k-l} \\
 &\quad - (2-\gamma)(1-\gamma)^{k-l}] \\
 &\leq \frac{\delta + D_\eta}{(\gamma - \phi) \mu} \left(\frac{2-\phi}{\phi} - \frac{2-\gamma}{\gamma} \right) \sum_{k=0}^{T-1} \frac{\alpha_k}{\theta_{\zeta,k+1}} \\
 &\leq \frac{\delta + D_\eta}{\mu \gamma \phi} D_{\alpha,\zeta} < \infty, \quad \forall T \geq 0.
 \end{aligned} \tag{36}$$

From Algorithm 1, recall that we fixed the observation sequence, the probability comes from the noise ξ_k and ζ_k . Therefore, the probability of execution is reduced to

$$\mathbb{P}[\mathcal{R}^{-1}(\mathcal{P}, \mathcal{O}, \mathbf{s}_0, \tilde{\mathbf{w}}_0, \mathbf{w}_0)] = \iint f_{\xi\zeta}(\xi, \zeta) d\xi d\zeta,$$

where $\iint f_{\xi\zeta}(\xi, \zeta) d\xi d\zeta = \prod_{i=1}^N \prod_{k=0}^{\infty} f_L(\xi_{i,k}, \theta_{\xi_i,k}) f_L(\zeta_{i,k}, \theta_{\zeta_i,k})$. According to (27), (28), (35) and (36), we derive

$$\begin{aligned}
 &\frac{\mathbb{P}[\mathcal{R}^{-1}(\mathcal{P}, \mathcal{O}, \mathbf{s}_0, \tilde{\mathbf{w}}_0, \mathbf{w}_0)]}{\mathbb{P}[\mathcal{R}^{-1}(\mathcal{P}', \mathcal{O}, \mathbf{s}_0, \tilde{\mathbf{w}}_0, \mathbf{w}_0)]} \\
 &= \prod_{k=0}^{\infty} \frac{f_L(\xi_{i_0,k}, \theta_{\xi_{i_0,k}}) f_L(\zeta_{i_0,k}, \theta_{\zeta_{i_0,k}})}{f_L(\xi'_{i_0,k}, \theta_{\xi'_{i_0,k}}) f_L(\zeta'_{i_0,k}, \theta_{\zeta'_{i_0,k}})} \\
 &\leq \prod_{k=0}^{\infty} e^{\frac{\|\Delta \xi_{i_0,k}\|_1}{\theta_{\xi,k}} + \frac{\|\Delta \zeta_{i_0,k}\|_1}{\theta_{\zeta,k}}} \\
 &= \exp \sum_{k=0}^{\infty} \left(\frac{\|\Delta s_{i_0,k}\|_1}{\theta_{\xi,k}} + \frac{\|\Delta \tilde{w}_{i_0,k}\|_1}{\theta_{\zeta,k}} \right) \\
 &\leq \exp \left(\frac{\delta + D_\eta}{\mu \gamma \phi} (D_{\alpha,\xi} + \phi D_{\alpha,\zeta}) \right).
 \end{aligned}$$

Therefore, we have $\epsilon = \frac{\delta + D_\eta}{\mu \gamma \phi} (D_{\alpha,\xi} + \phi D_{\alpha,\zeta})$.

From Theorem 3, we observe that the privacy level ϵ is proportional to $D_{\alpha,\xi} = \sum_{k=0}^{\infty} \frac{\alpha_k}{\theta_{\xi,k}}$ and $D_{\alpha,\zeta} = \sum_{k=0}^{\infty} \frac{\alpha_k}{\theta_{\zeta,k}}$. Note that α_k reflects the mismatch between supply and demand. To enhance the privacy of DP-DGT, we can reduce step sizes or increase noise power, but this comes at the cost of accuracy. Therefore, there is a trade-off between privacy and convergence accuracy.

We summarize the theoretical results in Theorem 2 and 3 and conclude that it is possible to choose the parameters α_k , $\theta_{\xi,k}$ and $\theta_{\zeta,k}$ such that DP-DGT can lead $\{\mathbf{w}_k\}$ to converge to a neighborhood of \mathbf{w}^* almost surely while achieving ϵ -differential privacy.

Corollary 1. Consider DP-DGT under Assumptions 1–4. When $\sum_{k=0}^{\infty} \alpha_k < \infty$, $\sum_{k=0}^{\infty} \theta_{\xi,k}^2 < \infty$, $\sum_{k=0}^{\infty} \theta_{\zeta,k}^2 < \infty$, $\sum_{k=0}^{\infty} \frac{\alpha_k}{\theta_{\xi,k}} < \infty$,

$\sum_{k=0}^{\infty} \frac{\alpha_k}{\theta_{\zeta,k}} < \infty$, $\sum_{k=0}^{\infty} \frac{\theta_{\xi,k}^2}{\alpha_k} < \infty$, $\sum_{k=0}^{\infty} \frac{\theta_{\zeta,k}^2}{\alpha_k} < \infty$, and there exists λ satisfying $q_C < \lambda < 1$ and $q_R < \lambda < 1$ and $k_0 > 0$ such that $\frac{\alpha_k}{\alpha_{k_0}} \geq \beta \lambda^{k-k_0}$, then $\{\mathbf{w}_k\}$ converges to a neighborhood of \mathbf{w}^* almost surely and while achieving ϵ -differential privacy simultaneously.

There indeed exists a possibility of choosing α_k , $\theta_{\xi,k}$ and $\theta_{\zeta,k}$ such that all conditions listed in Corollary 1 can be satisfied simultaneously. For example, we can let α_k , $\theta_{\xi,k}$ and $\theta_{\zeta,k}$ decrease linearly and further derive a close form of expression of ϵ .

Corollary 2. Consider DP-DGT under Assumptions 1–4. Let $\alpha_k = \alpha_0 q^k$, $\theta_{\xi,k} = \theta_{\xi,0} q_\xi^k$, and $\theta_{\zeta,k} = \theta_{\zeta,0} q_\zeta^k$. If $\alpha_0 < \mu \gamma \phi$, and $\{q_R, q_C, q_\xi^2, q_\zeta^2\} < q < \{q_\xi, q_\zeta\} < 1$, then

$$\epsilon = \frac{\alpha_0 \delta (\gamma \phi \mu + \alpha_0)}{\gamma \phi \mu (\gamma \phi \mu - \alpha_0)} \left(\frac{q_\xi}{\theta_{\xi,0} (q_\xi - q)} + \frac{\phi q_\zeta}{\theta_{\zeta,0} (q_\zeta - q)} \right), \tag{37}$$

with the error bound

$$\begin{aligned}
 &\mathbb{E}[\|\mathbf{w}_k - \mathbf{w}^*\|^2] \\
 &\leq O \left(1 + \frac{1}{(q - q_\zeta^2)(q - q_\xi^2)(q - q_C)(q - q_R)(1 - q)} \right).
 \end{aligned} \tag{38}$$

Proof. The proof is provided in Appendix A.4 (Huo et al., 2024).

As shown in previous works using δ -adjacency (Ding et al., 2021a, 2021b), the privacy loss is proportional to δ in (37). The value of δ reflects the distance between adjacent problems, indicating that larger differences require more noise for privacy preservation. Additionally, the result in (38) highlights the trade-off between noise level and convergence accuracy.

Remark 5. In contrast to Chen et al. (2023), where cumulative privacy loss increases indefinitely, the cumulative privacy loss ϵ in (37) remains constant even as the number of iterations grows. Additionally, unlike (Chen et al., 2023) and Huang et al. (2015), we derive this result without assuming bounded gradients.

6. Numerical simulations

The future microgrid is evolving into a cyber-physical system with three layers (Huang et al., 2010): a physical layer (power network), a communication layer (information transmission network), and a control layer (running distributed algorithms and processing communication data). Each bus node has a corresponding agent, and they exchange information via the communication network. We consider an economic dispatch problem for the IEEE 14-bus power system, where the generator buses are $\{1, 2, 3, 6, 8\}$ and the load buses are $\{2, 3, 4, 5, 6, 9, 10, 11, 13, 14\}$. Notably, the communication network among buses can be independent of the actual bus connections (Yang et al., 2013). We model the directed communication network as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where \mathcal{V} is the set combining the generators buses and load buses, and $\mathcal{E} = \{(i, i+1), (i, i+2) | 1 \leq i \leq 12\} \cup \{(13, 14), (13, 1), (14, 1), (1, 7), (2, 8), (3, 2), (3, 9), (4, 10), (5, 2), (5, 11), (6, 12)\}$. The cost functions of the generator i is $F_i(w_i) = a_i w_i^2 + b_i w_i + c_i$. The generator parameters, including the parameters of the quadratic cost functions, are adapted from (Kar & Hug, 2012) and presented in Table 2. When a bus does not contain generators, the power generation at that bus is set to zero. Thus, the update in (12b) simply becomes $w_{i,k} = 0$ for $i \notin \{1, 2, 3, 6, 8\}$. The virtual local demands at each bus are given as $D_1 = 0$ MW, $D_2 = 9$ MW, $D_3 = 56$ MW, $D_4 = 55$ MW, $D_5 = 27$ MW, $D_6 = 27$ MW, $D_7 = 0$ MW, $D_8 = 0$ MW, $D_9 = 8$ MW, $D_{10} = 24$ MW, $D_{11} = 53$ MW, $D_{12} = 46$ MW, $D_{13} = 16$ MW, and $D_{14} = 40$ MW. The total demand is $D = \sum_{i=1}^{14} D_i = 361$ MW, which is unknown to the agent at each bus. The optimal solution \mathbf{w}^* is obtained by using the CVX solver in a centralized manner, which is $[76.7398, 85.6530, 59.1311, 68.9863, 70.4898]^T$.

Table 2
IEEE 14-bus system generator parameters.

Bus	a_i (MW ² h)	b_i (\$/MWh)	Range (MW)
1	0.04	2.0	[0, 80]
2	0.03	3.0	[0, 90]
3	0.035	4.0	[0, 70]
6	0.03	4.0	[0, 70]
8	0.04	2.5	[0, 80]

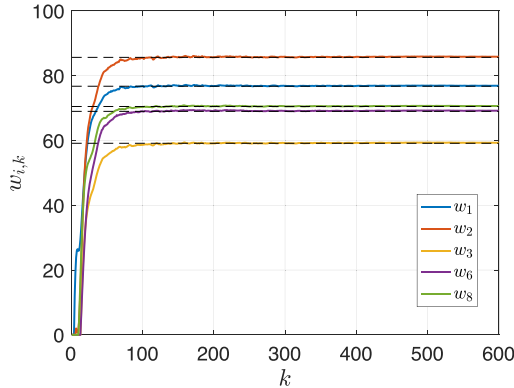


Fig. 2. Convergence performance of the generated power from Generators 1, 2, 3, 6, and 8.

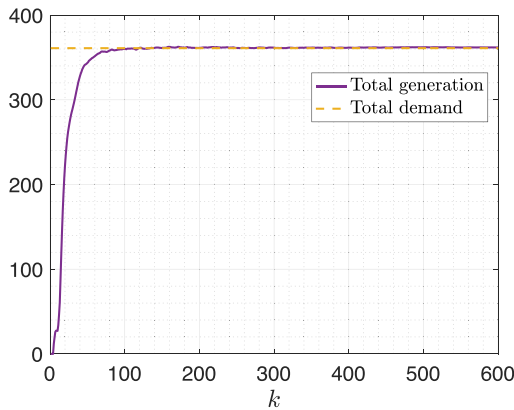


Fig. 3. Total generated power and demand.

6.1. Convergence of DP-DGT

We set the step size parameters to $\alpha_0 = 0.015$ and $q = 0.991$. The Laplacian noise are chosen as $\theta_{\xi,0} = \theta_{\zeta,0} = 0.01$ and $q_{\xi} = q_{\zeta} = 0.995$. Additionally, we select $\gamma = 0.8$ and $\phi = 0.7$. Fig. 2 shows that the decisions of generator buses 1, 2, 3, 6, and 8 converge to small neighborhoods of the optimal allocation solutions. Fig. 3 shows that the total generation asymptotically converges to a neighborhood of the total demand.

6.2. Tradeoff between accuracy and privacy

To demonstrate the tradeoff between convergence accuracy and the privacy level. We let $\theta_{\xi,0} = \theta_{\zeta,0} = \theta_0$, fix $\alpha_0 = 0.015$, $q = 0.991$, $q_{\xi} = q_{\zeta} = 0.995$, and vary θ_0 from 0 to 0.1. Due to the randomness of the Laplacian noise, we run the simulation 2000 times and obtain the empirical mean. Fig. 4 plots $\mathbb{E}[\|\mathbf{w}_{\infty} - \mathbf{w}^*\|^2]$ and $\frac{1}{\theta_0}$ under different intense of Laplacian noise, where the latter represents the trend of ϵ . Roughly speaking, Fig. 4 shows that as θ_0 increases, the expected convergence error $\mathbb{E}[\|\mathbf{w}_{\infty} - \mathbf{w}^*\|^2]$

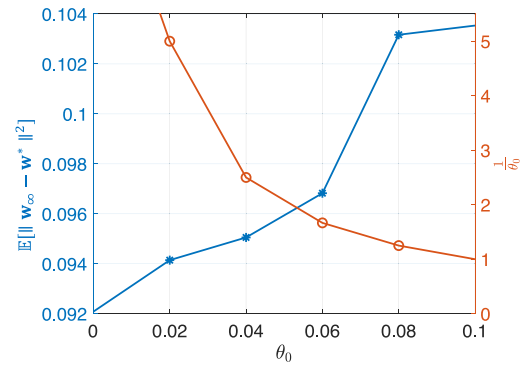


Fig. 4. Variations of $\mathbb{E}[\|\mathbf{w}_{\infty} - \mathbf{w}^*\|^2]$ and the trend of ϵ with different noise intensity θ_0 .

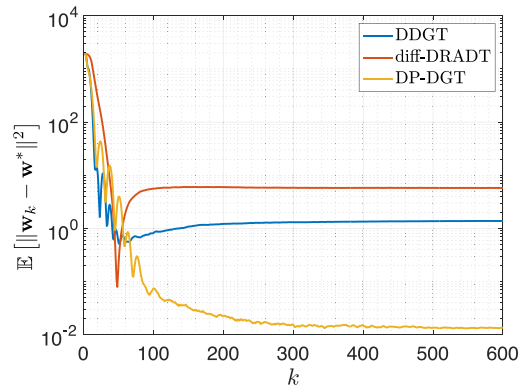


Fig. 5. Comparison of our proposed algorithm with the conventional DDGT algorithm (Zhang et al., 2020) with the same privacy level and the diff-DRADT in Ding et al. (2021a).

increases. Moreover, as θ_0 increases, ϵ increases as well. Thus, the tradeoff between the privacy level and the convergence is illustrated.

6.3. Comparison with state-of-the-art

We further compare the proposed algorithm with the conventional distributed dual gradient tracking algorithm (DDGT) expressed as (9) in Zhang et al. (2020) and the differentially private distributed resource allocation via deviation tracking algorithm (diff-DRADT) in Ding et al. (2021a). The step size and the noise is set to $\alpha_k = 0.034 \times 0.99^k$ and $\theta_{\xi,k} = \theta_{\zeta,k} = 0.01 \times 0.995^k$, respectively. Since there is no privacy preservation in conventional DDGT, to be fair, we run it with the same noise parameter and let $\iota\beta_k = \alpha_k = 0.034 \times 0.99^k$. Specifically, we set $\beta_k = 0.99^k$ and $\iota = 0.034$. Ding et al. (2021a) used the constant step size and linearly decreasing noise to achieve the finite cumulative privacy loss. Hence, we set the noise the same as ours and set the step size as 0.015. Fig. 5 depicts the comparison results. Since diff-DRADT only considers undirected graphs, it suffers from a high optimization error in the directed graph. Due to the lack of robustness, directly adding noise in DDGT will cause noise accumulation and accuracy compromise. It can be seen that under the same noise, our algorithm achieves the best convergence accuracy.

7. Conclusion and future work

This paper investigates privacy preservation in distributed resource allocation problems over directed unbalanced networks.

We propose a novel differentially private distributed deviation tracking algorithm that incorporates noises into transmitted messages to ensure privacy. The distribution of noise and step sizes are carefully designed to guarantee convergence and achieve ϵ -differential privacy simultaneously.

Potential future research directions include exploring methods to relax the step size requirement and accelerate convergence. Another interesting topic is to deal with the tradeoff between convergence accuracy and privacy level.

References

- Beaude, O., Benchimol, P., Gaubert, S., Jacquot, P., & Oudjane, N. (2020). A privacy-preserving method to optimize distributed resource allocation. *SIAM Journal on Optimization*, 30(3), 2303–2336.
- Bertsekas, D. P. (1997). Nonlinear programming. *Journal of the Operational Research Society*, 48(3), 334–334.
- Boyd, S. P., & Vandenberghe, L. (2004). *Convex optimization*. Cambridge University Press.
- Chen, W.-N., Choo, C. A. C., Kairouz, P., & Suresh, A. T. (2022). The fundamental price of secure aggregation in differentially private federated learning. In *International conference on machine learning* (pp. 3056–3089).
- Chen, X., Huang, L., He, L., Dey, S., & Shi, L. (2023). A differentially private method for distributed optimization in directed networks via state decomposition. *IEEE Transactions on Control of Network Systems*.
- Ding, T., Zhu, S., Chen, C., Xu, J., & Guan, X. (2021). Differentially private distributed resource allocation via deviation tracking. *IEEE Transactions on Signal and Information Processing over Networks*, 7, 222–235.
- Ding, T., Zhu, S., He, J., Chen, C., & Guan, X. (2021). Differentially private distributed optimization via state and direction perturbation in multiagent systems. *IEEE Transactions on Automatic Control*, 67(2), 722–737.
- Dwork, C. (2006). Differential privacy. In *International colloquium on automata, languages, and programming* (pp. 1–12). Springer.
- Freris, N. M., & Patrinos, P. (2016). Distributed computing over encrypted data. In *54th annual allerton conference on communication, control, and computing* (pp. 1116–1122).
- Han, S., Topcu, U., & Pappas, G. J. (2016). Differentially private distributed constrained optimization. *IEEE Transactions on Automatic Control*, 62(1), 50–64.
- Horn, R. A., & Johnson, C. R. (2012). *Matrix analysis*. Cambridge University Press.
- Hu, J., Chen, G., Li, H., Huang, T., & Ran, L. (2023). Achieving linear convergence for differentially private full-decentralized economic dispatch over directed networks. *Information Sciences*, 642, Article 119199.
- Huang, A. Q., Crow, M. L., Heydt, G. T., Zheng, J. P., & Dale, S. J. (2010). The future renewable electric energy delivery and management (FREEDM) system: the energy internet. *Proceedings of the IEEE*, 99(1), 133–148.
- Huang, Z., Mitra, S., & Vaidya, N. (2015). Differentially private distributed optimization. In *Proceedings of the 16th international conference on distributed computing and networking* (pp. 1–10).
- Huang, L., Wu, J., Shi, D., Dey, S., & Shi, L. (2024). Differential privacy in distributed optimization with gradient tracking. *IEEE Transactions on Automatic Control*.
- Huo, W., Chen, X., Huang, L., Johansson, K. H., & Shi, L. (2024). Differentially private dual gradient tracking for distributed resource allocation. arXiv preprint arXiv:2403.18275.
- Kar, S., & Hug, G. (2012). Distributed robust economic dispatch in power systems: A consensus+ innovations approach. In *IEEE power and energy society general meeting* (pp. 1–8).
- Lü, Q., Deng, S., Li, H., & Huang, T. (2023). Privacy protection decentralized economic dispatch over directed networks with accurate convergence. *IEEE Transactions on Emerging Topics in Computational Intelligence*.
- Munoz, A., Syed, U., Vassilvtiskii, S., & Vitercik, E. (2021). Private optimization without constraint violations. In *International conference on artificial intelligence and statistics* (pp. 2557–2565). PMLR.
- Pu, S. (2020). A robust gradient tracking method for distributed optimization over directed networks. In *59th IEEE conference on decision and control* (pp. 2335–2341).
- Pu, S., Shi, W., Xu, J., & Nedić, A. (2020). Push-pull gradient methods for distributed optimization in networks. *IEEE Transactions on Automatic Control*, 66(1), 1–16.
- Tsianos, K. I., Lawlor, S., & Rabbat, M. G. (2012). Push-sum distributed dual averaging for convex optimization. In *IEEE 51st conference on decision and control* (pp. 5453–5458).
- Wang, Y., & Başar, T. (2023). Decentralized nonconvex optimization with guaranteed privacy and accuracy. *Automatica*, 150, Article 110858.
- Wang, Y., & Başar, T. (2023). Quantization enabled privacy protection in decentralized stochastic optimization. *IEEE Transactions on Automatic Control*, 68(7), 4038–4052.
- Wang, Y., & Nedić, A. (2023). Tailoring gradient methods for differentially-private distributed optimization. *IEEE Transactions on Automatic Control*.
- Wang, Y., & Nedić, A. (2024). Robust constrained consensus and inequality-constrained distributed optimization with guaranteed differential privacy and accurate convergence. *IEEE Transactions on Automatic Control*.
- Wang, Y., & Poor, H. V. (2022). Decentralized stochastic optimization with inherent privacy protection. *IEEE Transactions on Automatic Control*, 68(4), 2293–2308.
- Wood, A. J., Wollenberg, B. F., & Sheblé, G. B. (2013). *Power generation, operation, and control*. John Wiley & Sons.
- Xi, C., & Khan, U. A. (2017). DEXTRA: A fast algorithm for optimization over directed graphs. *IEEE Transactions on Automatic Control*, 62(10), 4980–4993.
- Xiao, L., Johansson, M., & Boyd, S. P. (2004). Simultaneous routing and resource allocation via dual decomposition. *IEEE Transactions on Communications*, 52(7), 1136–1144.
- Yang, T., Lu, J., Wu, D., Wu, J., Shi, G., Meng, Z., & Johansson, K. H. (2016). A distributed algorithm for economic dispatch over time-varying directed networks with delays. *IEEE Transactions on Industrial Electronics*, 64(6), 5095–5106.
- Yang, S., Tan, S., & Xu, J.-X. (2013). Consensus based approach for economic dispatch problem in a smart grid. *IEEE Transactions on Power Systems*, 28(4), 4416–4426.
- Yi, P., Hong, Y., & Liu, F. (2016). Initialization-free distributed algorithms for optimal resource allocation with feasibility constraints and application to economic dispatch of power systems. *Automatica*, 74, 259–269.
- Zhang, Z., & Chow, M.-Y. (2012). Convergence analysis of the incremental cost consensus algorithm under different communication network topologies in a smart grid. *IEEE Transactions on Power Systems*, 27(4), 1761–1768.
- Zhang, J., You, K., & Cai, K. (2020). Distributed dual gradient tracking for resource allocation in unbalanced networks. *IEEE Transactions on Signal Processing*, 68, 2186–2198.
- Zhou, X. (2018). On the fenchel duality between strong convexity and lipschitz continuous gradient. arXiv preprint arXiv:1803.06573.



Wei Huo received her B.S. degree in Electronic and Information Engineering from Huazhong University of Science and Technology, Wuhan, China, in 2020, and Ph.D. degree in Electrical and Computer Engineering from Hong Kong University of Science and Technology, Hong Kong, in 2024. From August 2023 to December 2023, she was a visiting student in the School of Electrical Engineering and Computer Science in KTH Royal Institute of Technology. In December 2024, she joined Wireless Technology Lab, Central Research Institute, 2012, at Huawei, Shanghai, China, where she is currently a Senior Engineer. Her research interests include multi-agent systems, distributed optimization, privacy protection, and agentic AI for 6G.



optimization.

Xiaomeng Chen received her B.S. degree in Electrical Science and Engineering from Nanjing University, Jiangsu, China, in 2019, and the Ph.D. degree in Electrical and Computer Engineering from Hong Kong University of Science and Technology, Hong Kong, in 2023. She is currently a Postdoctoral Fellow at the School of Electrical and Computer Engineering, Hong Kong University of Science and Technology, Hong Kong. Her current research interests include cyber-physical system security/privacy, compressed communication, event-triggered mechanism and distributed



Lingying Huang is currently a National Youth Talent, Distinguished Young Professor, and Ph.D./Master's Supervisor at Southeast University, where she joined in December 2024. She earned her Ph.D. in Control Science and Engineering from the Hong Kong University of Science and Technology (HKUST) and subsequently conducted postdoctoral research at Nanyang Technological University (NTU), Singapore. During her doctoral studies, she was honored with the HKUST School of Engineering Academic Award for Continuing Ph.D. Students 2019–20. Her research contributions have been published in leading international journals, including *Automatica*, *IEEE Transactions on Automatic Control* (TAC), *IEEE Transactions on Signal Processing* (TSP), and *IEEE/ACM Transactions on Networking* (TON). She actively serves as a

reviewer for prestigious journals such as *Automatica* and *IEEE TAC*. Dr. Huang's research focuses on privacy and security in information systems, event-driven energy scheduling and real-time reconfiguration, multi-agent cooperative control, and robust nonlinear multi-sensor control for underwater robots, reflecting her interdisciplinary expertise in control theory, signal processing, and intelligent systems.



Karl Henrik Johansson is Professor with the School of Electrical Engineering and Computer Science at KTH Royal Institute of Technology in Sweden and Director of Digital Futures. He received M.Sc. degree in Electrical Engineering and Ph.D. in Automatic Control from Lund University. He has held visiting positions at UC Berkeley, Caltech, NTU, HKUST Institute of Advanced Studies, and NTNU. His research interests are in networked control systems and cyber-physical systems with applications in transportation, energy, and automation networks. He is President of the European Control

Association and member of the IFAC Council, and has served on the IEEE Control Systems Society Board of Governors and the Swedish Scientific Council for Natural Sciences and Engineering Sciences. He has received several best paper awards and other distinctions from IEEE, IFAC, and ACM. He has been awarded Swedish Research Council Distinguished Professor, Wallenberg Scholar with the Knut and Alice Wallenberg Foundation, Future Research Leader Award from the Swedish Foundation for Strategic Research, the triennial IFAC Young Author Prize, and IEEE Control Systems Society Distinguished Lecturer. He is Fellow of the IEEE and the Royal Swedish Academy of Engineering Sciences.



Ling Shi received his B.E. degree in Electrical and Electronic Engineering from The Hong Kong University of Science and Technology (HKUST) in 2002 and Ph.D. degree in Control and Dynamical Systems from The California Institute of Technology (Caltech) in 2008. He is currently a Professor in the Department of Electronic and Computer Engineering at HKUST with a joint appointment in the Department of Chemical and Biological Engineering (2025–2028). His research interests include cyber-physical systems security, networked control systems, sensor scheduling, event-based state

estimation, and multi-agent robotic systems (UAVs and UGVs). He served as an editorial board member for the European Control Conference 2013–2016. He was a subject editor for *International Journal of Robust and Nonlinear Control* (2015–2017), an associate editor for *IEEE Transactions on Control of Network Systems* (2016–2020), an associate editor for *IEEE Control Systems Letters* (2017–2020), and an associate editor for a special issue on *Secure Control of Cyber Physical Systems* in *IEEE Transactions on Control of Network Systems* (2015–2017). He also served as the General Chair of the 23rd International Symposium on Mathematical Theory of Networks and Systems (MTNS 2018). He is currently serving as a member of the Engineering Panel (Joint Research Schemes) of the Hong Kong Research Grants Council (RGC) (2023–2026). He received the 2024 Chen Han-Fu Award given by the Technical Committee on Control Theory, Chinese Association of Automation (TCCT, CAA). He is a member of the Young Scientists Class 2020 of the World Economic Forum (WEF), a member of The Hong Kong Young Academy of Sciences (YASHK), and he is an IEEE Fellow.