

# Non-Ergodic Convergence Algorithms for Distributed Consensus and Coupling-Constrained Optimization

Chenyang Qiu, Zongli Lin

**Abstract**—We study distributed convex optimization with two ubiquitous forms of coupling: consensus constraints and global affine equalities. We first design a linearized method of multipliers for the consensus optimization problem. Without smoothness or strong convexity, we establish non-ergodic sublinear rates of order  $O(1/\sqrt{k})$  for both the objective optimality and the consensus violation. Leveraging duality, we then show that the economic dispatch problem admits a dual consensus formulation, and that applying the same algorithm to the dual economic dispatch yields non-ergodic  $O(1/\sqrt{k})$  decay for the error of the summation of the cost over the network and the equality-constraint residual under convexity and Slater’s condition. Numerical results on the IEEE 118-bus system demonstrate faster reduction of both objective error and feasibility error relative to the state-of-the-art baselines, while the dual variables reach network-wide consensus.

## I. INTRODUCTION

This paper studies large-scale convex optimization problems formulated over networks, which frequently arise in engineering applications. For example, problems such as large-scale machine learning [1], distributed control [2], and economic dispatch [3] can be formulated as convex programs in networked systems. Compared with centralized algorithms, distributed algorithms do not rely on coordination from a central node and eliminate potential communication bottlenecks in computational infrastructure, such as high latency or low bandwidth. Furthermore, networks without a central node offer inherent advantages in terms of privacy protection and scalability. A central challenge, however, is to maintain feasibility and reach optimality when each agent observes only local information and communication is limited.

Two broad forms of coupling arise. The first is consensus coupling, which ties agents. Representative approaches include consensus-based subgradient methods [4] [5], distributed gradient methods [6] [7], primal–dual method [8], and algorithms for composite optimization [9]–[11]. These algorithms update local decision variables using (sub)gradients of the node-wise objective functions and aggregate information according to the communication graph, with the goal that, after multiple iterations, all nodes reach consensus while the iterates converge to a stationary point (or a neighborhood thereof).

The other broad form of coupling is the global equality relation binding the agents. A classical equality-coupling instance is economic dispatch problem, whose power-balance constraint couples all generators. For quadratic costs, average-consensus techniques yield explicit distributed solutions [12], [13]. For general objective functions, [14] provided a class

of algorithms that solve the economic dispatch problem with convergence guarantees based on the alternating direction method of multipliers (ADMM). Furthermore, [15] proposed an another ADMM-based algorithm which gets rid of the central coordinator. Ref. [16] proposed an ADMM-based algorithm, called augmented Lagrangian tracking, for distributed optimization with affine equality and nonlinear inequality coupling constraints.

Based on the duality theory, some algorithms were designed by first formulating the economic dispatch problem’s dual problem, which is then solved by consensus optimization algorithms. Ref. [17] proposed the Mirror-P-EXTRA algorithm to solve the economic dispatch problem in a distributed manner and provided the non-ergodic convergence rate of Karush–Kuhn–Tucker (KKT) conditions. In [18], a distributed dual gradient-tracking method is proposed; in the absence of local constraints, this method achieves ergodic sublinear convergence of the KKT residual under strongly convex cost functions. Ref. [19] solved the economic dispatch problem over time-varying networks with asymptotic convergence.

When the objective is nonsmooth and non-strongly convex, many distributed methods provide only asymptotic convergence [16] or ergodic guarantees for the objective value, i.e., rates for the running average along the local iteration sequence [9], [10], [18]. Such ergodic bounds allow pronounced last-iterate oscillations in the objective, which can be undesirable for real-time operation and stability considerations. In addition, several works primarily quantify the decay of KKT/feasibility (consensus) residuals rather than the objective suboptimality itself [11], [17], [18]. Notably, the non-ergodic objective error is particularly interpretable in engineering applications, as it directly reflects the *cost gap* at each iteration.

This paper develops a distributed framework that unifies consensus optimization and economic dispatch. For consensus problems, we recast the consensus constraint as a linear equality (equivalently, a graph-Laplacian feasibility condition) and design a linearly augmented Lagrangian with fully distributed updates. Under the assumption of the convexity of the objective function alone, we prove that the sequence generated by the proposed algorithm can guarantee the non-ergodic sublinear convergence rate  $O(1/\sqrt{k})$  of the objective error and the consensus error. Leveraging the primal–dual relation between the consensus optimization and economic dispatch problem, we use the proposed algorithm to solve the economic dispatch problem. Under the assumption of the convexity of the objective function and the Slater’s condition, we demonstrate that the sequence generated by the proposed algorithm can guarantee the non-ergodic sublinear convergence rate  $O(1/\sqrt{k})$  simultaneously for the global cost and

(Corresponding author: Zongli Lin.)

Chenyang Qiu and Zongli Lin are with the Charles L. Brown Department of Electrical and Computer Engineering, University of Virginia, Charlottesville, VA 22904, USA (e-mail: nzp4an@virginia.edu; zl5y@virginia.edu).

the constraint violation.

The contributions of this paper are summarized as follows.

- 1) We develop a distributed method for consensus optimization with nonsmooth convex objective functions. With a constant stepsize, we provide the non-ergodic convergence rates for both objective optimality error and consensus error.
- 2) Based on the dual decomposition technique, we reformulate economic dispatch as a dual *consensus* problem and adapt the algorithm accordingly. Under the assumption of convexity and Slater's condition, we provide the non-ergodic convergence rates for both objective optimality error and constraint violation.
- 3) Numerical experiments conducted on the IEEE 118-bus system demonstrate that our algorithm converges faster than the state-of-the-art algorithms for both objective error and feasibility residuals.

The remainder of this paper is organized as follows. Section II develops an algorithm for consensus optimization and analyzes its convergence. Section III studies the economic dispatch problem. Section IV provides numerical experiments to validate our algorithm. Section V concludes the paper.

**Notation:** For a function  $g : \mathbb{R}^m \rightarrow \mathbb{R}$ , its subgradient set at  $x \in \mathbb{R}^m$  is represented by  $\partial g(x)$ , i.e.,  $\tilde{\nabla}g(x) \in \partial g(x)$  is a subgradient of  $g$  at  $x$ . A convex function  $f : X \rightarrow (-\infty, +\infty]$  is said to be proper if  $f(x) > -\infty$  for all  $x \in X$  and  $f(x)$  is not trivially equal to  $+\infty$ . We use the symbols  $1_p$ ,  $0_p$ ,  $O_p$ , and  $I_p$  to denote the  $p$ -dimensional all-one vector, all-zero vector, zero matrix, and identity matrix, respectively. Let  $\otimes$  denote the Kronecker product,  $\langle \cdot, \cdot \rangle$  denote the Euclidean inner product, and  $\|\cdot\|$  denote the  $\ell_2$  norm. We write  $\Omega \succeq O_p$  if it is positive semidefinite and  $\Omega \succ O_p$  if it is positive definite. For any  $\Omega \succeq O_p$  and  $x \in \mathbb{R}^p$ ,  $\|x\|_\Omega^2 := x^\top \Omega x$ ,  $\Omega^{\frac{1}{2}}$  is the square root of  $\Omega$  (i.e.,  $\Omega^{\frac{1}{2}} \Omega^{\frac{1}{2}} = \Omega$ ), and  $\Omega^\dagger$  is the pseudoinverse of  $\Omega$ . For a set  $C \subseteq \mathbb{R}^p$ , the corresponding indicator function is denoted by  $\mathbf{1}_C$ .

## II. CONSENSUS OPTIMIZATION

This section introduces the distributed optimization problem with consensus constraints and formulates it into a linear constrained compact form which can be solved by a linearized method of multipliers.

### A. Problem Formulation

In this paper, we consider a distributed system consisting of  $n$  agents that exchange information over a connected undirected communication network. The network is modeled by a graph  $(\mathcal{V}, \mathcal{E})$ , where  $\mathcal{V} = \{1, 2, \dots, n\}$  is the set of nodes and  $\mathcal{E} \subseteq \{(i, j) \subseteq \mathcal{V} \times \mathcal{V} \mid i \neq j\}$  is the set of edges. For each agent,  $i \in \mathcal{V}$ , we denote the set of its neighbors by  $\mathcal{N}_i = \{j \in \mathcal{V} \mid (i, j) \in \mathcal{E}\}$ . All the agents in this network cooperate to minimize the total cost by solving the following problem,

$$\begin{aligned} \min_{y_i \in Y_i} \quad & \sum_{i=1}^n g_i(y_i) \\ \text{s. t.} \quad & y_1 = y_2 = \dots = y_n, \end{aligned} \quad (1)$$

where  $Y_i \subseteq \mathbb{R}^p$ ,  $y_i$  is the local copy of the decision variable and  $f_i$  is the local objective function owned by each agent  $i \in \mathcal{V}$ . We denote the corresponding Laplacian matrix by  $\mathcal{L}$ , where

$$[\mathcal{L}]_{ij} = \begin{cases} \sum_{s \in \mathcal{N}_i} H_{is}, & i = j, \\ -H_{ij}, & j \in \mathcal{N}_i, \\ 0, & \text{otherwise,} \end{cases} \quad i, j \in \mathcal{V},$$

with  $\mathcal{L}_{ij} = \mathcal{L}_{ji} > 0$  being the weight of edge  $\{i, j\} \in \mathcal{E}$ . Since  $(\mathcal{V}, \mathcal{E})$  is a connected undirected graph, the null space of  $\mathcal{L}$  is  $\text{span}\{1_n\}$ . Define  $\mathcal{Y} = Y_1 \times Y_2 \times \dots \times Y_n$ ,  $y := [y_1^\top y_2^\top \dots y_n^\top]^\top \in \mathcal{Y}$ ,  $W := \mathcal{L} \otimes I_p \succeq O_{np}$ . Then, we obtain that the range spaces of the matrices  $W$ ,  $W^{\frac{1}{2}}$ ,  $W^\dagger$  and  $(W^\dagger)^{\frac{1}{2}}$  are the same and equal to  $\{y \in \mathbb{R}^{np} \mid y_1 + \dots + y_n = 0_p\}$ , which is the orthogonal complement of  $\{y \in \mathbb{R}^{np} \mid y_1 = \dots = y_n\}$ . Hence, the consensus constraint of  $y$  is equivalent to  $W^{\frac{1}{2}}y = 0_{np}$ . The optimal solution of problem (1), denoted by  $y^* = [(y_1^*)^\top (y_2^*)^\top \dots (y_n^*)^\top]^\top \in \mathcal{Y}$  must satisfy  $W^{\frac{1}{2}}y^* = 0_{np}$ . Problem (1) can therefore be written in the compact form

$$\begin{aligned} \min_{y \in \mathcal{Y}} \quad & G(y) = \sum_{i=1}^n g_i(y_i) \\ \text{s. t.} \quad & W^{\frac{1}{2}}y = 0_{np}. \end{aligned} \quad (2)$$

### B. Algorithm Development

**Assumption 1.** For each  $i \in \mathcal{V}$ , we assume that  $Y_i \subseteq \mathbb{R}^p$  is a convex set and  $g_i(x_i)$  is a proper and convex function.

Based on the above formulation and Assumption 1, we aim to develop a distributed algorithm for solving problem (2) by utilizing the linearized method of multipliers. To this end, we define the augmented Lagrangian function as

$$L_\rho(y, v) = G(y) - \left\langle v, W^{\frac{1}{2}}y \right\rangle + \frac{\rho}{2} \|W^{\frac{1}{2}}y\|^2,$$

where  $v = [v_1^\top v_2^\top \dots v_n^\top]^\top \in \mathbb{R}^{n(d+m)}$  and  $\rho > 0$ . Recall the method of multipliers [20], where updates are given by

$$y_{k+1} = \underset{y \in \mathcal{Y}}{\operatorname{argmin}} \left\{ G(y) - \left\langle v_k, W^{\frac{1}{2}}y \right\rangle + \frac{\rho}{2} \|W^{\frac{1}{2}}y\|^2 \right\}, \quad (3a)$$

$$v_{k+1} = v_k - \rho W^{\frac{1}{2}}y_{k+1}. \quad (3b)$$

Since the quadratic term  $\frac{\rho}{2} \|W^{\frac{1}{2}}y\|^2$  are not suitable for distributed computation, we linearize it as  $\rho \langle Wy_k, y \rangle + \frac{\eta}{2} \|y - y_k\|^2$ , where  $\eta \geq 0$  is a designed parameter for the algorithm. Besides, to avoid using  $W^{\frac{1}{2}}$ , we define

$$\lambda_k = W^{\frac{1}{2}}v_k \quad (4)$$

for distributed implementation, so that  $\lambda_{k+1} = \lambda_k - \rho W y_{k+1}$ . Therefore, we get the compact form of the proposed iteration for problem (1)

$$\begin{aligned} y_{k+1} = \underset{y \in \mathcal{Y}}{\operatorname{argmin}} \quad & \left\{ G(y) - \langle \lambda_k, y \rangle + \rho \langle Wy_k, y \rangle \right. \\ & \left. + \frac{\eta}{2} \|y - y_k\|^2 \right\} \end{aligned} \quad (5a)$$

$$\lambda_{k+1} = \lambda_k - \rho W y_{k+1}. \quad (5b)$$

The distributed implementation of the above algorithm over the undirected network  $(\mathcal{V}, \mathcal{E})$  is detailed in Algorithm 1.

---

**Algorithm 1**


---

- 1: **Initialization:** Each agent  $i \in \mathcal{V}$ , sets  $\lambda_{i,0}$  such that  $\sum_{i=1}^n \lambda_{i,0} = 0_p$  and arbitrarily sets  $y_{i,0} \in Y_i$  and. Each agent  $i \in \mathcal{V}$  sends the variable  $y_{i,0}$  to its neighbors  $j \in \mathcal{N}_i$ . After receiving the information from its neighbors, each agent  $i \in \mathcal{V}$  computes the aggregated information  $t_{i,0} = \sum_{j \in \mathcal{N}_i} p_{ij} (y_{i,0} - y_{j,0})$ .
  - 2: **for**  $k = 0, 1, 2, \dots$ , **do**
  - 3:   Each agent  $i \in \mathcal{V}$  updates the variable  $y_{i,k+1} = \operatorname{argmin}_{y_i \in Y_i} \{g_i(y_i) - \langle \lambda_{i,k} - \rho t_{i,k}, y_i \rangle + \frac{\eta}{2} \|y_i - y_{i,k}\|^2\}$ .
  - 4:   Each agent  $i \in \mathcal{V}$  sends the variable  $y_{i,k+1}$  to its neighbors  $j \in \mathcal{N}_i$ . After receiving the information from its neighbors, each agent  $i \in \mathcal{V}$  computes the aggregated information  $t_{i,k+1} = \sum_{j \in \mathcal{N}_i} p_{ij} (y_{i,k+1} - y_{j,k+1})$ .
  - 5:   Each agent  $i \in \mathcal{V}$  updates  $\lambda_{i,k+1} = \lambda_{i,k} - \beta_k t_{i,k+1}$ .
  - 6: **end for**
- 

### C. Convergence Analysis

We denote the variables generated by the proposed algorithm at the  $k$ th iteration by  $z_k = [y_k^\top \ \lambda_k^\top]^\top$  and the difference between the  $(k+1)$ th iteration and the  $k$ th iteration by  $\Delta y_{k+1} = y_k - y_{k+1}$ ,  $\Delta \lambda_{k+1} = \lambda_k - \lambda_{k+1}$ , and  $\Delta z_{k+1} = z_k - z_{k+1}$ .

**Lemma 1.** Suppose Assumption 1 holds. If  $\eta I - \rho W > 0$ , then the sequences  $y_k$  and  $\lambda_k$  generated by the proposed algorithm satisfy the following inequality,

$$\begin{aligned} & \frac{1}{k} \sum_{s=0}^k (\|\Delta y_{s+1}\|_\Omega^2 + \|\Delta \lambda_{s+1}\|_\Omega^2) \\ & \leq \frac{1}{k} (\|y_0 - y^*\|_\Omega^2 + \frac{1}{\rho} \|\lambda_0 - \lambda^*\|_{W^\dagger}^2), \end{aligned} \quad (6)$$

where  $\Omega = \eta I - \rho W$ .

*Proof.* According to (5a), we have

$$\begin{aligned} & \lambda_k - \rho W y_k - \eta(y_{k+1} - y_k) \\ & \in \partial G(y_{k+1}) + \partial \mathbf{1}_{\mathcal{Y}}(y_{k+1}), \end{aligned} \quad (7)$$

where  $\mathbf{1}_{\mathcal{Y}}$  is the indicator function of  $\mathcal{Y}$ ,  $\partial G(y_{k+1})$  and  $\partial \mathbf{1}_{\mathcal{Y}}(y_{k+1})$  are the sets of subgradients of  $G$  and  $\mathbf{1}_{\mathcal{Y}}$  at  $y_{k+1}$ , respectively. Based on KKT conditions [21], we have that any vector  $\lambda^*$  satisfying the following equation is a dual optimal point corresponding to problem (1)

$$\lambda^* \in \partial G(y^*) + \partial \mathbf{1}_{\mathcal{Y}}(y^*). \quad (8)$$

Therefore, plus with (5b), for any  $\tilde{\nabla}G(y_{k+1}) \in \partial G(y_{k+1})$  and  $\tilde{\nabla}\mathbf{1}_{\mathcal{Y}}(y_{k+1}) \in \partial \mathbf{1}_{\mathcal{Y}}(y_{k+1})$ , we have

$$\begin{aligned} & \tilde{\nabla}\mathbf{1}_{\mathcal{Y}}(y_{k+1}) - \tilde{\nabla}\mathbf{1}_{\mathcal{Y}}(y^*) + \Omega(y_{k+1} - y_k) - (\lambda_{k+1} - \lambda^*) \\ & = -\tilde{\nabla}G(y_{k+1}) + \tilde{\nabla}G(y^*), \end{aligned} \quad (9)$$

where  $\Omega = \eta I_{n(d+m)} - \rho W \succ O_{n(d+m)}$ . By the convexity of  $G$ , we have

$$0 \leq \langle \tilde{\nabla}G(y_{k+1}) - \tilde{\nabla}G(y^*), y_{k+1} - y^* \rangle$$

$$\begin{aligned} & = -\langle \tilde{\nabla}\mathbf{1}_{\mathcal{Y}}(y_{k+1}) - \tilde{\nabla}\mathbf{1}_{\mathcal{Y}}(y^*), y_{k+1} - y^* \rangle \\ & \quad - \langle \Omega(y_{k+1} - y_k), y_{k+1} - y^* \rangle \\ & \quad + \langle \lambda_{k+1} - \lambda^*, y_{k+1} - y^* \rangle \\ & \leq -\langle \Omega(y_{k+1} - y_k), y_{k+1} - y^* \rangle \\ & \quad + \langle \lambda_{k+1} - \lambda^*, y_{k+1} - y^* \rangle \\ & = -\langle \Omega(y_{k+1} - y_k), y_{k+1} - y^* \rangle \\ & \quad - \left\langle \lambda_{k+1} - \lambda^*, \frac{W^\dagger}{\rho} (\lambda_{k+1} - \lambda_k) \right\rangle \\ & = \frac{1}{2} \left( \|y_k - y^*\|_\Omega^2 - \|y_{k+1} - y_k\|_\Omega^2 - \|y_{k+1} - y^*\|_\Omega^2 \right) \\ & \quad + \frac{1}{2\rho} \left( \|\lambda_k - \lambda^*\|_{W^\dagger}^2 - \|\lambda_{k+1} - \lambda_k\|_{W^\dagger}^2 \right. \\ & \quad \left. - \|\lambda_{k+1} - \lambda^*\|_{W^\dagger}^2 \right), \end{aligned} \quad (10)$$

which indicates that  $\forall k > 1$ ,  $\|y_k - y^*\|_\Omega^2 + \|\lambda_k - \lambda^*\|_{W^\dagger}^2$  is nonincreasing, i.e.,

$$\begin{aligned} & \|y_k - y^*\|_\Omega^2 + \|\lambda_k - \lambda^*\|_{W^\dagger}^2 \\ & \geq \|y_{k+1} - y^*\|_\Omega^2 + \frac{1}{\rho} \|\lambda_{k+1} - \lambda^*\|_{W^\dagger}^2. \end{aligned} \quad (11)$$

Sum (10) from  $s = 1$  to  $s = k$  yields

$$\begin{aligned} & \frac{1}{k} \sum_{s=0}^k (\|\Delta y_{s+1}\|_\Omega^2 + \|\Delta \lambda_{s+1}\|_\Omega^2) \\ & \leq \frac{1}{k} (\|y_0 - y^*\|_\Omega^2 + \frac{1}{\rho} \|\lambda_0 - \lambda^*\|_{W^\dagger}^2), \end{aligned} \quad (12)$$

and thus the proof of Lemma 1 is completed.  $\square$

**Lemma 2.** Suppose the conditions of Lemma 1 hold. For  $k > 0$ ,

$$\|\Delta z_{k+1}\|_{\hat{\Omega}} \leq \frac{1}{\sqrt{k}} (\|z_0 - z^*\|_{\hat{\Omega}}), \quad (13)$$

where  $\hat{\Omega} = \begin{bmatrix} \Omega & O_{np} \\ O_{np} & \frac{1}{\rho} W^\dagger \end{bmatrix}$ .

*Proof.* Applying (9) for  $k+1$  and  $k$ , we have

$$\begin{aligned} & -\Omega(\Delta y_{k+1} - \Delta y_k) + \Delta \lambda_{k+1} \\ & = \tilde{\nabla}G(y_{k+1}) - \tilde{\nabla}G(y_k) + \tilde{\nabla}\mathbf{1}_{\mathcal{Y}}(y_{k+1}) - \tilde{\nabla}\mathbf{1}_{\mathcal{Y}}(y_k) \end{aligned} \quad (14)$$

By (5b) and the convexity of  $G$  and  $\mathbf{1}_{\mathcal{Y}}$ , we have

$$\begin{aligned} 0 & \leq \langle \tilde{\nabla}G(y_{k+1}) - \tilde{\nabla}G(y_k) + \tilde{\nabla}\mathbf{1}_{\mathcal{Y}}(y_{k+1}) - \tilde{\nabla}\mathbf{1}_{\mathcal{Y}}(y_k), \\ & \quad y_{k+1} - y_k \rangle \\ & = \langle -\Omega(\Delta y_{k+1} - \Delta y_k) + \Delta \lambda_{k+1}, \Delta y_{k+1} \rangle \\ & = \langle -\Omega(\Delta y_{k+1} - \Delta y_k), \Delta y_{k+1} \rangle \\ & \quad - \frac{1}{\rho} \langle \Delta \lambda_{k+1}, W^\dagger (\Delta \lambda_{k+1} - \Delta \lambda_k) \rangle \\ & = -\frac{1}{2} (\|\Delta y_{k+1}\|_\Omega^2 - \|\Delta y_k\|_\Omega^2 - \|\Delta y_{k+1} - \Delta y_k\|_\Omega^2) \\ & \quad - \frac{1}{2\rho} (\|\Delta \lambda_{k+1}\|_{W^\dagger}^2 - \|\Delta \lambda_k\|_{W^\dagger}^2 \\ & \quad - \|\Delta \lambda_{k+1} - \Delta \lambda_k\|_{W^\dagger}^2), \end{aligned} \quad (15)$$

which indicates that  $\Delta z_{k+1}$  is nonincreasing, i.e.,

$$\|\Delta z_{k+1}\|_{\hat{\Omega}}^2 \leq \|\Delta z_k\|_{\hat{\Omega}}^2. \quad (16)$$

Combining (6) and (16) yields (13) in Lemma 2.  $\square$

To characterize the convergence rates of the optimality gap and the consensus error, we establish the following theorem.

**Theorem 1.** Suppose the conditions of Lemma 1 hold, then the sequence  $z_k$  generated by the proposed algorithm satisfies

$$\begin{aligned} & \frac{1}{\sqrt{k}} \left( \frac{1}{2} \|z_0 - z^*\|_{\hat{\Omega}}^2 + \|\tilde{\nabla}G(y^*)\| \|z_0 - z^*\|_{\hat{\Omega}} \right) \\ & \geq G(y_{k+1}) - G(y^*) \\ & \geq - \frac{1}{\sqrt{k}} \|\tilde{\nabla}G(y^*)\| \|z_0 - z^*\|_{\hat{\Omega}}, \end{aligned} \quad (17)$$

and

$$\|y_{k+1}\|_W = \frac{1}{\sqrt{k}} (\|z_0 - z^*\|_{\hat{\Omega}}). \quad (18)$$

*Proof.* By the convexity of  $G$ , we have

$$\begin{aligned} & G(y_{k+1}) - G(y^*) \\ & \leq \langle \tilde{\nabla}G(y_{k+1}) - \tilde{\nabla}G(y^*), y_{k+1} - y^* \rangle + \langle \tilde{\nabla}G(y^*), y_{k+1} - y^* \rangle \end{aligned} \quad (19)$$

According to (10), the first term of the left-hand side of (19) is bounded by the following inequalities

$$\begin{aligned} & \langle \tilde{\nabla}G(y_{k+1}) - \tilde{\nabla}G(y^*), y_{k+1} - y^* \rangle \\ & \leq \frac{1}{2} \left( \|y_k - y^*\|_{\Omega}^2 - \|y_{k+1} - y_k\|_{\Omega}^2 - \|y_{k+1} - y^*\|_{\Omega}^2 \right) \\ & \quad + \frac{1}{2\rho} \left( \|\lambda_k - \lambda^*\|_{W^\dagger}^2 - \|\lambda_{k+1} - \lambda_k\|_{W^\dagger}^2 \right. \\ & \quad \left. - \|\lambda_{k+1} - \lambda^*\|_{W^\dagger}^2 \right) \\ & = \frac{1}{2} \left( \langle \Omega(y_k - y_{k+1}), y_k + y_{k+1} - y^* \rangle - \|y_{k+1} - y_k\|_{\Omega}^2 \right) \\ & \quad + \frac{1}{2\rho} \left( \langle W^\dagger(\lambda_k - \lambda_{k+1}), \lambda_k + \lambda_{k+1} - 2\lambda^* \rangle \right. \\ & \quad \left. - \|\lambda_{k+1} - \lambda_k\|_{W^\dagger}^2 \right) \\ & \leq \frac{1}{4} \left( \|y_k - y_{k+1}\|_{\Omega} (\|y_k - y^*\|_{\Omega} + \|y_{k+1} - y^*\|_{\Omega}) \right. \\ & \quad \left. - \|y_{k+1} - y_k\|_{\Omega}^2 \right) \\ & \quad + \frac{1}{4\rho} \left( \|\lambda_k - \lambda_{k+1}\|_{W^\dagger} (\|\lambda_k - \lambda^*\|_{W^\dagger} + \|\lambda_{k+1} - \lambda^*\|_{W^\dagger}) \right. \\ & \quad \left. - \|\lambda_{k+1} - \lambda_k\|_{W^\dagger}^2 \right) \\ & \leq \frac{\|z_0 - z^*\|_{\hat{\Omega}}}{4\sqrt{k}} \left( (\|y_k - y^*\|_{\Omega} + \|y_{k+1} - y^*\|_{\Omega}) \right. \\ & \quad \left. + (\|\lambda_k - \lambda^*\|_{W^\dagger} + \|\lambda_{k+1} - \lambda^*\|_{W^\dagger}) \right) \\ & \leq \frac{\|z_0 - z^*\|_{\hat{\Omega}}^2}{2\sqrt{k}}. \end{aligned} \quad (20)$$

As for the second term of the right-hand side of (19), we have

$$\begin{aligned} \langle \tilde{\nabla}G(y^*), y_{k+1} - y^* \rangle & \leq \langle \tilde{\nabla}G(y^*), y_{k+1} \rangle \\ & \leq \frac{1}{\rho} \|\tilde{\nabla}G(y^*)\| \|\lambda_{k+1} - \lambda_k\|_{W^\dagger} \\ & \leq \frac{\|\tilde{\nabla}G(y^*)\|}{\sqrt{k}} (\|z_0 - z^*\|_{\hat{\Omega}}). \end{aligned} \quad (21)$$

where the first inequality is due to (5b) and the fact that  $y^*$  is in the null space of  $\lambda_{k+1}$ . Combining (20) and (21) yields the first inequality of (17). Similarly, for the lower bound of the optimality gap in (17),

$$\begin{aligned} G(y_{k+1}) - G(y^*) & \geq \langle \tilde{\nabla}G(y^*), y_{k+1} - y^* \rangle \\ & \geq - \|\tilde{\nabla}G(y^*)\| \|\lambda_{k+1} - \lambda_k\|_{W^\dagger} \\ & \geq - \frac{\|\tilde{\nabla}G(y^*)\|}{\sqrt{k}} (\|z_0 - z^*\|_{\hat{\Omega}}). \end{aligned} \quad (22)$$

$\square$

### III. EXTENSION TO ECONOMIC DISPATCH

In this section, we study the economic dispatch problem, establish its primal–dual correspondence with consensus optimization, and, leveraging this link, adapt Algorithm 1 to it and provide the corresponding convergence analysis.

#### A. Problem Formulation

The objective of the economic dispatch problem is to minimize the total cost of power generation in the network while meeting the total power demand and satisfying individual constraints. For each generator  $i \in \mathcal{V}$ , let  $x_i$  be its power output and  $f_i(x_i)$  be its local cost function. Let  $d$  represent the total demand that must be satisfied by the power output of all generators. Initially, each generator  $i$  is assigned a virtual local demand  $d_i$ , with  $\sum_{i=1}^n d_i = \bar{d}$ . Then, the economic dispatch problem is formulated as a constrained optimization problem, which has the following general form,

$$\begin{aligned} \min_x f(x) &= \sum_{i=1}^n f_i(x_i), \\ \text{s.t. } \sum_{i=1}^n x_i &= \sum_{i=1}^n d_i = \bar{d}, \quad x \in X, \end{aligned} \quad (23)$$

where  $x = [x_1^T \ x_2^T \ \dots \ x_n^T]^T \in \mathbb{R}^{np}$ ,  $X_i \subseteq \mathbb{R}^p$  is the local constraint set for each  $x_i$ , and  $X = X_1 \times X_2 \times \dots \times X_n$ . Note that  $p = 1$  is usually set in the field of power systems. However, we will address the constrained optimization problem (23) where  $p$  is allowed to be any finite positive integer.

**Assumption 2.** For each  $i \in \mathcal{V}$ , we assume that  $f_i(x_i) : \mathbb{R}^p \rightarrow (-\infty, +\infty]$  is a proper and convex function.

**Assumption 3.** The set  $X_i$  is nonempty, closed, and convex. Furthermore, the Slater's condition is satisfied, i.e., the constraint  $\sum_{i=1}^n x_i - d_i = 0$  is satisfied for at least one point in the relative interior of  $X$ .

#### B. Primal-Dual Relationship

By introducing the Lagrangian multiplier  $\delta \in \mathbb{R}^p$ , the Lagrangian function associated with problem (23) is defined as  $L(x, \delta) = f(x) + \langle \delta, \sum_{i=1}^n (x_i - d_i) \rangle$ . The dual problem to problem (23) is then given as

$$\begin{aligned} \max_{\delta \in \mathbb{R}^p} \min_x L(x, \delta) &= \max_{\delta \in \mathbb{R}^p} \sum_{i=1}^n \min_{x_i \in X_i} \{f_i(x_i) + \langle \delta, x_i - d_i \rangle\} \\ &= \max_{\delta \in \mathbb{R}^p} \sum_{i=1}^n -f_i^*(-\delta) - \langle \delta, d_i \rangle, \end{aligned} \quad (24)$$

where for each  $i \in \mathcal{V}$ , the convex conjugate function  $f_i^*(\delta)$  is defined as

$$\begin{aligned} f_i^*(\delta) &= \max_{x_i \in X_i} \{\langle \delta, x_i \rangle - f_i(x_i)\} \\ &= - \min_{x_i \in X_i} \{f_i(x_i) - \langle \delta, x_i \rangle\}. \end{aligned}$$

Let  $g_i(\delta) = f_i^*(-\delta) + \langle \delta, d_i \rangle$ . Then, the dual problem in (24) becomes

$$\min_{\delta \in \mathbb{R}^p} \sum_{i=1}^n g_i(\delta).$$

Assign a local copy of the dual variable  $y_i$  to each agent  $i \in \mathcal{V}$ , we have the consensus optimization problem

$$\begin{aligned} \min_{y \in \mathbb{R}^{np}} \quad G(y) &= \sum_{i=1}^n g_i(y_i) \\ \text{s. t.} \quad y_1 &= y_2 = \dots = y_n, \end{aligned} \tag{25}$$

which is equivalent to (1).

Denote a optimal solution of problem (23) by  $x^* = [(x_1^*)^\top \ (x_2^*)^\top \ \dots \ (x_n^*)^\top]^\top$ . Then,  $x^*$  and  $y^*$  as the optimal pair of the Lagrangian function  $L(x, y)$  if and only if

- 1)  $x^*$  is the optimal solution to problem (23), i.e.,  $f(x^*) \leq f(x)$  for any  $x \in X$  and  $\sum_{i=1}^n x_i^* = d$ .
- 2)  $y^*$  is an optimal solution to the dual problem (25).

Assumptions 2 and 3 guarantee the strong duality holds [22], i.e.,  $G(y^*) = -f(x^*)$ .

### C. Algorithm Development

With the form of the consensus optimization problem, we can use iteration (5) to solve the dual problem of (23). Define  $d = [d_1^\top \ d_2^\top \ \dots \ d_n^\top]^\top$ . Applying the definition of  $G$  into (5) gives the following iteration,

$$\begin{aligned} y_{k+1} &= \operatorname{argmin}_{y \in \mathbb{R}^{np}} \left\{ G(y) - \langle \lambda_k, y \rangle + \rho \langle W y_k, y \rangle + \frac{\eta}{2} \|y - y_k\|^2 \right\} \\ &= \operatorname{argmin}_{y \in \mathbb{R}^{np}} \max_{x \in X} \left\{ -(f(x) - \langle x - d, y \rangle) - \langle \lambda_k - \rho W y_k, y \rangle \right. \\ &\quad \left. + \frac{\eta}{2} \|y - y_k\|^2 \right\} \\ &= y_k + \frac{1}{\eta} (\tilde{x} - d + \lambda_k - \rho W y_k), \end{aligned} \tag{26}$$

where  $\tilde{x}$  is the optimal value solved by

$$\tilde{x} = \max_{x \in X} \left\{ -(f(x) - \langle x - d, y_{k+1} \rangle) - \langle \lambda_k - \rho W y_k, y_{k+1} \rangle \right. \\ \left. + \frac{\eta}{2} \|y_{k+1} - y_k\|^2 \right\} \tag{27}$$

Plugging (26) into (27) yields

$$\begin{aligned} x_{k+1} &= \operatorname{argmin}_{x \in X} \left\{ f(x) + \frac{\eta}{2} \|y_k + \frac{1}{\eta} (x - d + \lambda_k - \rho W y_k)\|^2 \right\}, \end{aligned} \tag{28a}$$

$$\begin{aligned} y_{k+1} &= y_k + \frac{1}{\eta} (x_{k+1} - d + \lambda_k - \rho W y_k), \end{aligned} \tag{28b}$$

Since the objective function of the above subproblem is separable, the update of  $x_{k+1}$ ,  $y_{k+1}$ , and  $\lambda_{k+1}$  can be carried out in a distributed manner, which is described in Algorithm 2, where for each  $i \in \mathcal{V}$ ,  $J_{i,k}(\xi) = y_{i,k} + \frac{1}{\eta} (\xi - d_i + \lambda_{i,k} - t_{i,k})$ , and  $t_{i,k} = \sum_{j \in \mathcal{N}_i} p_{ij} (y_{i,k} - y_{j,k})$ .

---

### Algorithm 2

- 1: **Initialization:** Each agent  $i \in \mathcal{V}$ , sets  $\lambda_{i,0}$  such that  $\sum_{i=1}^n \lambda_{i,0} = 0_p$  and arbitrarily sets  $y_{i,0} \in \mathbb{R}^p$  and. Each agent  $i \in \mathcal{V}$  sends the variable  $y_{i,0}$  to its neighbors  $j \in \mathcal{N}_i$ . After receiving the information from its neighbors, each agent  $i \in \mathcal{V}$  computes the aggregated information  $t_{i,0} = \sum_{j \in \mathcal{N}_i} p_{ij} (y_{i,0} - y_{j,0})$ .
  - 2: **for**  $k = 0, 1, \dots$ , **do**
  - 3:   Each agent  $i \in \mathcal{V}$  updates the variable  $x_{i,k+1}$  by  $\operatorname{argmin}_{\xi \in X_i} \{f_i(\xi) + \frac{\eta}{2} \|J_{i,k}(\xi)\|^2\}$ .
  - 4:   Each agent  $i \in \mathcal{V}$  updates the variable  $y_{i,k+1} = J_{i,k}(x_{i,k+1})$  and then sends the variable  $y_{i,k+1}$  to its neighbors  $j \in \mathcal{N}_i$ .
  - 5:   After receiving the information from its neighbors, each agent  $i \in \mathcal{V}$  computes the aggregated information  $t_{i,k} = \sum_{j \in \mathcal{N}_i} p_{ij} (y_{i,k} - y_{j,k})$ , and further update  $\lambda_{i,k+1} = \lambda_{i,k} - \beta_k t_{i,k+1}$ .
  - 6: **end for**
- 

### D. Convergence Analysis

In this section, we present the convergence analysis for algorithm 2. Specifically, we provide the convergence rates of the optimality gap in the function value and the feasibility error.

**Theorem 2.** Consider the distributed algorithm in Algorithm 2 under Assumptions 2 and 3. If the parameters  $\eta$  and  $\rho$  of Algorithm 1 satisfies  $\eta I - \rho W \succ O_{n(d+m)}$ , then the difference between the summation of the cost over the network and the optimal cost at the  $(k+1)$ th iteration is bounded as follows,

$$\begin{aligned} &\frac{1}{\sqrt{k}} \left( \frac{1}{2} \|z_0 - z^*\|_{\hat{\Omega}}^2 + \|y^*\|_{\Omega} \|z_0 - z^*\|_{\hat{\Omega}} \right) \\ &\geq f(x_{k+1}) - f(x^*) \\ &\geq - \frac{1}{\sqrt{k}} \|y^*\|_{\Omega} \|z_0 - z^*\|_{\hat{\Omega}}. \end{aligned} \tag{29}$$

*Proof.* Taking the derivative of the function of the subproblem (28a) gives

$$\begin{aligned} &y_k + \frac{1}{\eta} (x_{k+1} - d + \lambda_k - \rho W y_k) \\ &\in -(\partial f(x_{k+1}) + \partial \mathbf{1}_{\mathcal{X}}(x_{k+1})), \end{aligned}$$

substitution of which in (28b) gives

$$y_{k+1} \in -(\partial f(x_{k+1}) + \partial \mathbf{1}_{\mathcal{X}}(x_{k+1})),$$

and similarly,

$$y^* \in -(\partial f(x^*) + \partial \mathbf{1}_{\mathcal{X}}(x^*)).$$

By the convexity of  $f$ , we have  $\square$

$$\begin{aligned} & f(x_{k+1}) - f(x^*) \\ & \leq \langle x^* - x_{k+1}, y_{k+1} - y^* \rangle + \langle x^* - x_{k+1}, y^* \rangle. \end{aligned} \quad (30)$$

Based on the definition of  $g_i$  and the conjugate duality,

$$\begin{aligned} -x_{i,k+1} + d_i & \in \partial g_i(y_{i,k+1}) \\ & = \partial f_i^*(-y_{i,k+1}) + d_i \\ & = -\arg \min_{x \in X_i} \{f_i(x) + \langle x, y_{i,k+1} \rangle\} + d_i, \end{aligned}$$

and similarly,

$$-x^* + d_i \in \partial g_i(y^*).$$

Therefore, by substituting the subgradients of  $G$  at  $y_{i,k+1}$  and  $y^*$  with  $-x_{i,k+1} + d_i$  and  $-x^* + d_i$  in (20), respectively, the first term of the right-hand side of (30) is upper bounded as follows,

$$\begin{aligned} & \langle x^* - x_{k+1}, y_{k+1} - y^* \rangle \\ & \leq \frac{1}{2} \left( \|y_k - y^*\|_\Omega^2 - \|y_{k+1} - y_k\|_\Omega^2 - \|y_{k+1} - y^*\|_\Omega^2 \right) \\ & \quad + \frac{1}{2\rho} \left( \|\lambda_k - \lambda^*\|_{W^\dagger}^2 - \|\lambda_{k+1} - \lambda_k\|_{W^\dagger}^2 \right. \\ & \quad \left. - \|\lambda_{k+1} - \lambda^*\|_{W^\dagger}^2 \right) \\ & = \frac{1}{2} \left( \langle \Omega(y_k - y_{k+1}), y_k + y_{k+1} - y^* \rangle - \|y_{k+1} - y_k\|_\Omega^2 \right) \\ & \quad + \frac{1}{2\rho} \left( \langle W^\dagger(\lambda_k - \lambda_{k+1}), \lambda_k + \lambda_{k+1} - 2\lambda^* \rangle \right. \\ & \quad \left. - \|\lambda_{k+1} - \lambda_k\|_{W^\dagger}^2 \right) \\ & \leq \frac{1}{4} \left( \|y_k - y_{k+1}\|_\Omega (\|y_k - y^*\|_\Omega + \|y_{k+1} - y^*\|_\Omega) \right. \\ & \quad \left. - \|y_{k+1} - y_k\|_\Omega^2 \right) \\ & \quad + \frac{1}{4\rho} \left( \|\lambda_k - \lambda_{k+1}\|_{W^\dagger} (\|\lambda_k - \lambda^*\|_{W^\dagger} + \|\lambda_{k+1} - \lambda^*\|_{W^\dagger}) \right. \\ & \quad \left. - \|\lambda_{k+1} - \lambda_k\|_{W^\dagger}^2 \right) \\ & \leq \frac{\|z_0 - z^*\|_{\hat{\Omega}}^2}{2\sqrt{k}}. \end{aligned} \quad (31)$$

Besides, since for any  $k \geq 0$ ,  $\lambda^*$  and  $\lambda_{k+1}$  are in the range space of  $W$  while  $y^*$  is in the null space of  $W$ , we have

$$\langle \lambda_{k+1} - \lambda^*, y^* \rangle = 0. \quad (32)$$

Based on (32), we have the following inequality

$$\begin{aligned} & \langle h(x^*) - h(x_{k+1}), y^* \rangle \\ & = \langle \eta(-\epsilon_{k+1}, y^*) + \langle \Omega(y_k - y_{k+1}), y^* \rangle + \langle \lambda_{k+1} - \lambda^*, y^* \rangle \rangle \\ & \leq \langle \Omega y_k - y_{k+1}, y^* \rangle \\ & = \|y^*\|_\Omega \|y_k - y_{k+1}\|_\Omega. \end{aligned} \quad (33)$$

Combining (31) and (33) yields the first inequality of (29). In the end, we get the second inequality of (29) by

$$\begin{aligned} & f(x_{k+1}) - f(x^*) \\ & \geq \langle h(x^*) - h(x_{k+1}), y^* \rangle \\ & \geq -\|y^*\|_\Omega \|y_k - y_{k+1}\|_\Omega. \end{aligned} \quad (34)$$

**Theorem 3.** Under the same condition of Theorem 2, we have the convergence rate of the feasibility error as follows,

$$\left\| \sum_{i=1}^n x_{i,k+1} - d_i \right\| \leq \frac{1}{\sqrt{k}} \|z_0 - z^*\|_{\hat{\Omega}}. \quad (35)$$

*Proof.*

$$\begin{aligned} & (1_n^T \otimes I_{d+m})(x_{k+1} - d) \\ & = (1_n^T \otimes I_{d+m})(\Omega(y_{k+1} - y_k) - (\lambda_{k+1} - \lambda^*)) \\ & = (1_n^T \otimes I_{d+m})\Omega(y_{k+1} - y_k), \end{aligned} \quad (36)$$

where the last equality is due to that  $\lambda_{k+1}$  and  $\lambda^*$  are in the range space of  $W$ . Therefore, we have

$$\begin{aligned} & \left\| \sum_{i=1}^n x_{i,k+1} - d_i \right\| \\ & = \|(1_n^T \otimes I_{d+m})(x_{k+1} - d)\| \\ & \leq \|(1_n^T \otimes I_{d+m})\| \|x_{k+1} - d\| \\ & \leq \|(1_n^T \otimes I_{d+m})\| \|\Omega\|^{\frac{1}{2}} \|y_{k+1} - y_k\|_\Omega \\ & \leq (n\|\Omega\|)^{\frac{1}{2}} \|y_{k+1} - y_k\|_\Omega \\ & \leq \frac{1}{\sqrt{k}} \|z_0 - z^*\|_{\hat{\Omega}}, \end{aligned} \quad (37)$$

which completes the proof of Theorem 3.  $\square$

#### IV. NUMERICAL EXPERIMENTS

In this section, we conduct several numerical experiments to validate the proposed algorithm and compare it with three existing state-of-the-art algorithms. Specifically, the proposed algorithm is compared with those in [18], [17] and [16], which are termed distributed dual gradient tracking algorithm (DDGT), Mirror-P-EXTRA and Augmented Lagrangian Tracking (ALT), respectively.

To illustrate the effectiveness of the proposed algorithm, we employ the IEEE 118-bus system [23]. The cyber network for the IEEE 118-bus system is constructed as a connected undirected graph  $(\mathcal{V}, \mathcal{E})$ , where  $\mathcal{V} = \{1, 2, \dots, 118\}$  represents the set of nodes (corresponding to buses) and the edge set  $\mathcal{E}$  is defined by the rule  $(i, i+1)$  and  $(i, i+2)$  for all  $1 \leq i \leq 116$ . This means that each node is connected to its nearest neighbor and the next-nearest neighbor, forming a sparse yet connected topology. Based on this topology, we further derive a randomly generated doubly stochastic weight matrix. In this system, generator buses are randomly set to be located at 14 different buses. In the economic dispatch problem, we set the local cost of each generator as a quadratic function  $a_i x_i^2 + b_i x_i + c_i$ . The coefficients for the local cost functions are adopted from [23]. We set  $[\underline{p}_i, \bar{p}_i] = [0, 300]$  for each generator. For buses without generators, their corresponding coefficients  $a_i$  and  $b_i$  are set to zero and their local constraints are set as  $\underline{p}_i = \bar{p}_i = 0$ . The total power demand across the system is given by  $\sum_{i=1}^{14} d_i = 950$ MW, where  $d$  is unknown to each bus. Without loss of generality, we set a virtual initial local demand  $d_i$  as  $(950/14)$ MW at each bus with a generator and 0MW at all other buses.

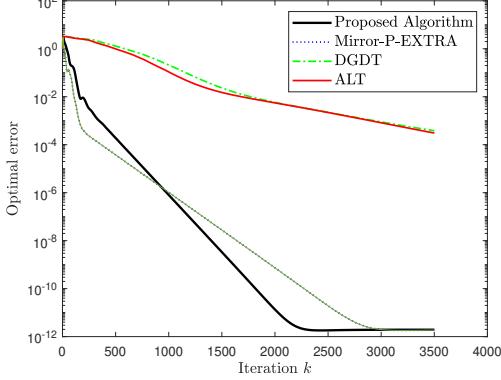


Fig. 1: The power mismatch under the five algorithms.

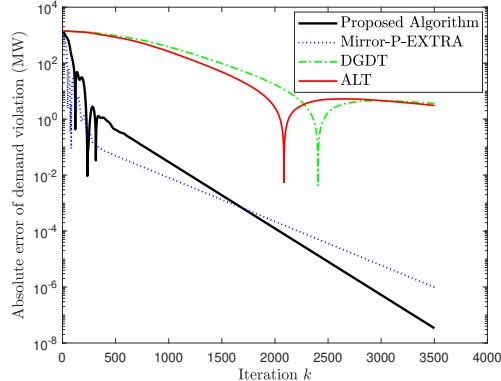


Fig. 2: The absolute error of the total power generation and the total power demand under the five algorithms.

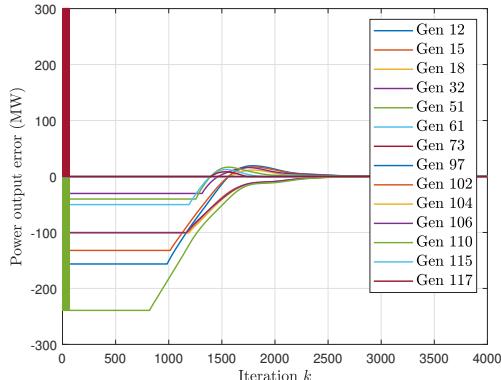


Fig. 3: The power output of each generator under the proposed algorithm.

Figure 1 plots the total optimal error of the cost function of the IEEE 118-bus system given the total power demand  $d$ , i.e.,  $\frac{|f(x_k) - f(x^*)|}{|f(x_1) - f(x^*)|}$  both for the proposed algorithm and for DDGT, Mirror-P-EXTRA, and ALT algorithms. As shown in Fig. 1, the convergence of the proposed algorithm is faster than that of all three other algorithms. Since the existence of strong duality, Fig. 1 also demonstrates the effectiveness of Algorithm 1 for the consensus optimization problem. Figure 2

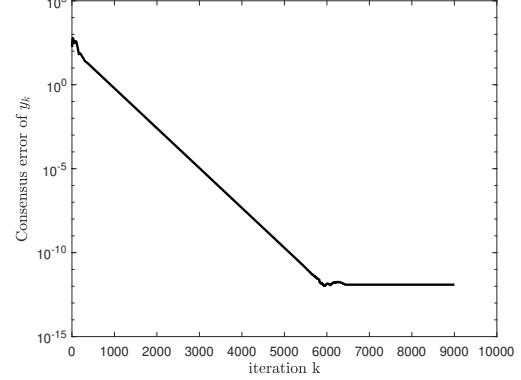


Fig. 4: The consensus error of the variable  $y_k$  under the proposed algorithm.

presents the absolute error of the total power generation and the total power demand, i.e.,  $|\sum_{i=1}^n (x_{i,k} - d_i)|$ . As seen, our algorithm again outperforms the others, reaching a violation error below  $10^{-6}$  earlier than all other algorithms.

Figure 3 plots the error between the power output of each generator and its optimal power output under the proposed algorithm, i.e.,  $|x_{i,k} - x_i^*|$ ,  $i \in \mathcal{V}$ . It can be seen in Fig. 3 that the error of each generator can gradually reach zero. In addition, Fig. 4 shows convergence of the consensus error  $\|y_k - 1_{118}^T y_k\|$  of the variable  $y_k$ .

## V. CONCLUSION

We presented a distributed framework for convex optimization over networks with consensus and global equality couplings. For consensus problems, a linearized augmented Lagrangian scheme achieves non-ergodic  $O(1/\sqrt{k})$  rates for both objective error and consensus violation. Via a dual consensus reformulation, the same method applied to economic dispatch yields non-ergodic  $O(1/\sqrt{k})$  decay of the objective error and the equality-constraint residual under convexity and Slater's condition. Experiments on the IEEE 118-bus system confirm faster reduction of objective and feasibility errors than representative state-of-the-art baselines.

## REFERENCES

- [1] J. Dean, G. Corrado, R. Monga, K. Chen, M. Devin, M. Mao, M. Ranzato, A. Senior, P. Tucker, K. Yang *et al.*, "Large scale distributed deep networks," *Advances in neural information processing systems*, vol. 25, 2012.
- [2] J. Zhao and F. Dörfler, "Distributed control and optimization in dc microgrids," *Automatica*, vol. 61, pp. 18–26, 2015.
- [3] F. Marzbani and A. Abdelfatah, "Economic dispatch optimization strategies and problem formulation: A comprehensive review," *Energies*, vol. 17, no. 3, p. 550, 2024.
- [4] A. Nedic, A. Ozdaglar, and P. A. Parrilo, "Constrained consensus and optimization in multi-agent networks," *IEEE Transactions on Automatic Control*, vol. 55, no. 4, pp. 922–938, 2010.
- [5] K. I. Tsianos and M. G. Rabbat, "Distributed strongly convex optimization," in 2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton). IEEE, 2012, pp. 593–600.
- [6] W. Shi, Q. Ling, G. Wu, and W. Yin, "Extra: An exact first-order algorithm for decentralized consensus optimization," *SIAM Journal on Optimization*, vol. 25, no. 2, pp. 944–966, 2015.

- [7] R. Xin and U. A. Khan, "A linear algorithm for optimization over directed graphs with geometric convergence," *IEEE Control Systems Letters*, vol. 2, no. 3, pp. 315–320, 2018.
- [8] P. Bianchi, W. Hachem, and F. Iutzeler, "A coordinate descent primal-dual algorithm and application to distributed asynchronous optimization," *IEEE Transactions on Automatic Control*, vol. 61, no. 10, pp. 2947–2957, 2015.
- [9] X. Wu and J. Lu, "A unifying approximate method of multipliers for distributed composite optimization," *IEEE Transactions on Automatic Control*, vol. 68, no. 4, pp. 2154–2169, 2022.
- [10] J. Xu, Y. Tian, Y. Sun, and G. Scutari, "Distributed algorithms for composite optimization: Unified framework and convergence analysis," *IEEE Transactions on Signal Processing*, vol. 69, pp. 3555–3570, 2021.
- [11] W. Shi, Q. Ling, G. Wu, and W. Yin, "A proximal gradient algorithm for decentralized composite optimization," *IEEE Transactions on Signal Processing*, vol. 63, no. 22, pp. 6013–6023, 2015.
- [12] R. Wang, Q. Li, B. Zhang, and L. Wang, "Distributed consensus based algorithm for economic dispatch in a microgrid," *IEEE Transactions on Smart Grid*, vol. 10, no. 4, pp. 3630–3640, 2019.
- [13] R. Wang, Q. Li, and H. Liu, "Distributed economic dispatch under random time-varying digraph," *IEEE Transactions on Industrial Informatics*, 2024.
- [14] T.-H. Chang, M. Hong, and X. Wang, "Multi-agent distributed optimization via inexact consensus ADMM," *IEEE Transactions on Signal Processing*, vol. 63, no. 2, pp. 482–497, 2014.
- [15] G. Chen and Q. Yang, "An ADMM-based distributed algorithm for economic dispatch in islanded microgrids," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 9, pp. 3892–3903, 2017.
- [16] A. Falsone and M. Prandini, "Augmented Lagrangian tracking for distributed optimization with equality and inequality coupling constraints," *Automatica*, vol. 157, p. 111269, 2023.
- [17] A. Nedić, A. Olshevsky, and W. Shi, "Improved convergence rates for distributed resource allocation," in *Proceedings of the IEEE Conference on Decision and Control (CDC)*, 2018, pp. 172–177.
- [18] J. Zhang, K. You, and K. Cai, "Distributed dual gradient tracking for resource allocation in unbalanced networks," *IEEE Transactions on Signal Processing*, vol. 68, pp. 2186–2198, 2020.
- [19] T. Yang, J. Lu, D. Wu, J. Wu, G. Shi, Z. Meng, and K. H. Johansson, "A distributed algorithm for economic dispatch over time-varying directed networks with delays," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 6, pp. 5095–5106, 2016.
- [20] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.
- [21] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [22] D. P. Bertsekas, "Nonlinear programming," *Journal of the Operational Research Society*, vol. 48, no. 3, pp. 334–334, 1997.
- [23] P. Venkatesh, R. Gnanadass, and N. P. Padhy, "Comparison and application of evolutionary programming techniques to combined economic emission dispatch with line flow constraints," *IEEE Transactions on Power Systems*, vol. 18, no. 2, pp. 688–697, 2003.