

A Distributed Gradient-based Algorithm for Optimization Problems with Coupled Equality Constraints [★]

Chenyang Qiu ^a, Zongli Lin ^a

^aCharles L. Brown Department of Electrical and Computer Engineering, University of Virginia, Charlottesville, VA 22904, USA

Abstract

This paper studies a class of distributed optimization problems with coupled equality constraints in networked systems. Many existing distributed algorithms rely on solving local subproblems via the argmin operator in each iteration. Such approaches become computationally burdensome or intractable when local cost functions are complex. To address this challenge, we propose a novel distributed gradient-based algorithm that avoids solving a local optimization problem at each iteration by leveraging first-order approximations and projection onto local feasible sets. The algorithm operates in a fully distributed manner, requiring only local communication without exchanging gradients or primal variables. We rigorously establish sublinear convergence for general convex cost functions and linear convergence under strong convexity and smoothness conditions. Numerical simulation on the IEEE 118-bus system demonstrates the superior computational efficiency and scalability of the proposed method compared to several state-of-the-art distributed optimization algorithms.

Key words: Distributed optimization, first-order method, coupled-constraint optimization

1 Introduction

The optimization problem with coupled equality constraints, also known as the resource allocation or economic dispatch problem, serves as a fundamental model in various real-world applications, including efficient power system operation [1], channel assignment in wireless communication networks [2], and task scheduling [3]. At its core, the optimization problems with coupled equality constraints seek to optimally allocate available resources to satisfy demand at the minimum possible cost, while adhering to a range of operational constraints. Historically, such problems have been solved via centralized approaches, where a central coordinator, equipped with global system information, computes the optimal solution [4]. While effective, these centralized methods incur significant communication overhead, raise privacy concerns, and are vulnerable to the failure of the central node. In the meantime, the rapid development of several advanced technologies, such as distributed energy resources, cyber-networks, and cloud

computing, is making conventional centralized management strategies increasingly inadequate [5, 6].

These practical limitations of centralized methods, together with the growing deployment of advanced technologies, have spurred the development of distributed algorithms. In the distributed framework, each agent conducts local optimization and communicates with neighboring units through limited information exchange. This decentralized strategy significantly improves scalability, enhances system resilience, and offers better privacy preservation. Consensus-based algorithms have been extensively adopted to solve the optimization problem with coupled equality constraints in a distributed manner. These algorithms exploit local communication to iteratively achieve consensus on marginal costs, enabling fully distributed and scalable solutions without the need for centralized coordination [7–10]. However, existing consensus-based approaches often assume quadratic cost functions and relatively simple constraints, which limit their applicability to more general settings involving non-quadratic objectives.

[★] This work was supported in part by the National Science Foundation under the grant number CMMI-2243930. The material in this paper was not presented at any conference. Corresponding author: Zongli Lin.

Email addresses: nzp4an@virginia.edu (Chenyang Qiu), z15y@virginia.edu (Zongli Lin).

To handle more general or nonsmooth cost functions beyond quadratic functions, a range of advanced distributed optimization algorithms has been proposed. A common approach involves reformulating the distributed optimization problem with coupled equality

constraints by introducing two convex sets with associated indicator functions, thereby enabling variable splitting and problem decomposition. Building on the Alternating Direction Method of Multipliers (ADMM) framework, several works have made notable progress [11–13]. Ref. [11] provided a class of algorithms that solve the distributed optimization problem with coupled equality constraints with convergence guarantees. Further, Ref. [12] introduced a fully distributed ADMM-based algorithm with convergence guarantees. Ref. [14] proposes a distributed primal–dual algorithm, where each iteration embeds a finite-time multi-step consensus iterations to get local estimates of the global constraint violation. More recently, Ref. [15] proposed an augmented Lagrangian tracking method to address distributed optimization problems involving coupled constraints, broadening the applicability of ADMM-based strategies.

Another line of research tries to solve the distributed optimization problem with coupled equality constraints from a duality perspective, by reformulating it as a consensus optimization problem and solving its dual formulation [16–20]. For instance, based on the mirror relationship between the consensus optimization problem and the optimization problem with coupled equality constraints, Ref. [16] proposed three algorithms, Mirror-P-EXTRA, Mirror-EXTRA, and Mirror-PG-EXTRA. Ref. [17] utilized the first-order information of the objective function and provided convergence rates for both convex and strongly convex cases. Leveraging the duality, Ref. [18] developed a distributed dual gradient tracking algorithm for resource allocation. Under the assumption of strong convexity and in the absence of local constraints, the algorithm achieves ergodic sublinear convergence to the Karush–Kuhn–Tucker (KKT) condition at a rate of $O(1/k)$. Algorithms proposed in Ref. [19] achieve an ergodic convergence rate of $O(1/k)$ considering local constraints and have been extended for differential privacy.

It is worth noting that many existing distributed algorithms for solving distributed optimization problems with coupled equality constraints rely heavily on the use of the argmin operator and require each agent to solve a local optimization problem at every iteration. While this approach can be efficient when local objective functions are simple, such as in the case of quadratic functions with closed-form solutions, its applicability becomes limited when dealing with more complex functions, such as exponential type cost functions [21], where solving the local optimization problem at each iteration is computationally expensive or even intractable.

To address this challenge, we propose a novel distributed algorithm, termed the Distributed Gradient-based Algorithm (DGA), that circumvents the need to solve local optimization problems at each iteration, resulting in a *single-loop* algorithm. Instead of applying the

argmin operator, DGA uses a first-order approximation of each agent’s objective function based solely on local gradient information, thereby significantly reducing the per-iteration computational burden. To ensure feasibility, the proposed algorithm employs projection onto local constraint sets at each iteration, maintaining variable updates within the allowable operational limits. Another appealing feature of our algorithm is that it does not require exchanging sensitive information, such as the gradient and primal variables, making it privacy-preserving. Under mild and standard assumptions, we rigorously establish the convergence properties of the algorithm. Specifically, when the cost function is convex and Lipschitz-smooth, we establish a nonergodic convergence rate of $o(1/k)$. Under the additional assumption of strong convexity, the convergence further improves to a linear rate. Numerical simulation demonstrates that our algorithm outperforms state-of-the-art methods in terms of convergence speed, both with respect to the number of iterations and the overall computation time.

Our main contributions in this paper are summarized as follows.

- (1) We propose a novel gradient-based distributed optimization algorithm for solving optimization problems with coupled affine equality constraints by adapting an approximate method of multipliers. In contrast to many existing approaches [11, 12, 14, 15, 18, 19, 22–24] that require solving a local optimization problem at each iteration, the proposed method eliminates the need for the use of argmin to solve local optimization problem per iteration. Instead, it relies solely on local gradient information and projection onto feasible sets, thereby significantly reducing per-iteration computational complexity. Unlike existing methods, the proposed algorithm does not require any special initialization and guarantees the feasibility of each agent’s local constraints throughout the iterations.
- (2) For convex cost functions, we prove that the KKT errors converge to 0 at a nonergodic rate of $o(1/k)$. For strongly convex functions, we establish linear convergence. In contrast, existing algorithms based on similar first-order approximations, such as Mirror-PG-EXTRA [16], which lacks explicit convergence guarantees, and DuSPA [17], which achieves a non-ergodic rate of $O(1/k)$ for convex case, offer weaker theoretical assurances.

The remainder of this paper is organized as follows. Section 2 introduces the problem formulation. Section 3 develops our single-loop distributed algorithm. Section 4 establishes the convergence analysis. Section 5 provides numerical experiments to validate our algorithm. Section 6 concludes the paper.

Notation: For a differentiable function $f : \mathbb{R}^p \rightarrow \mathbb{R}$, its gradient at $x \in \mathbb{R}^p$ is denoted by $\nabla f(x)$. For any,

possibly non-differentiable, function $g : \mathbb{R}^p \rightarrow \mathbb{R}$, its subdifferential, i.e., its subgradient set at $x \in \mathbb{R}^p$, is represented by $\partial g(x)$. The symbols 0_p , 1_p , and I_p are used to denote the p -dimensional all-zero vector, all-one vector, and identity matrix, respectively. In addition, $O_{m \times p}$ denotes the $(m \times p)$ -dimensional zero matrix. Let \otimes denote the Kronecker product, $\langle \cdot, \cdot \rangle$ be the Euclidean inner product, and $\|\cdot\|$ be the ℓ_2 norm. Given $A \in \mathbb{R}^{p \times p}$, $A^{\frac{1}{2}}$ is the square root of A (i.e., $A^{\frac{1}{2}}A^{\frac{1}{2}} = A$), and we write $A \succeq O_{p \times p}$ if it is positive semidefinite and $A \succ O_{p \times p}$ if it is positive definite. For any $A \succeq O_{p \times p}$ and $x \in \mathbb{R}^p$, $\|x\|_A^2 := x^T A x$, the spectral norm of A is denoted by $\|A\|$. For any matrix A , A^\dagger is A' 's pseudoinverse. In addition, we denote the block diagonal matrix by $\text{diag}(A_1, A_2, \dots, A_n)$ whose blocks are A_1, A_2, \dots, A_n . The projection of a vector x onto X is denoted by $\mathcal{P}_X(x)$.

2 Problem Formulation

2.1 Optimization Problem

This paper considers the following affine coupled-constraint problem,

$$\min_{x \in X} f(x) = \sum_{i=1}^n f_i(x_i), \quad \text{s. t. } \sum_{i=1}^n A_i x_i = \sum_{i=1}^n d_i, \quad (1)$$

where $x = [x_1^T \ x_2^T \ \dots \ x_n^T]^T \in \mathbb{R}^{np}$, $x_i \in \mathbb{R}^p$, $i \in \mathcal{V}$, are the decision variables, $d_i \in \mathbb{R}^m$, $A_i \in \mathbb{R}^{m \times p}$ and $\{A_i\}_{i=1}^n$ are not all zero, the local constraint for each agent is a convex constraint set $X_i \subseteq \mathbb{R}^p$. Accordingly, we define $X = X_1 \times X_2 \times \dots \times X_n$. In problem (1), we only require $m \leq np$, which is weaker than [19], where A_i is required to have a full row rank.

Assumption 1 For each $i \in \mathcal{V}$, the function f_i is convex, differentiable and l_f -smooth, i.e.,

$$\|\nabla f_i(x'_i) - \nabla f_i(x_i)\| \leq l_f \|x'_i - x_i\| \quad \forall x'_i, x_i \in \mathbb{R}^p,$$

for some $l_f > 0$.

Assumption 2 (Slater's condition) The solution set is nonempty, and there exist some points in the relative interior of X , $\text{ri}\{X\}$, such that the desired power constraint $\sum_{i=1}^n A_i x_i = \sum_{i=1}^n d_i$ is satisfied.

2.2 Communication Model

We consider a networked system consisting of n agents labeled by the index set $\mathcal{V} = \{1, 2, \dots, n\}$, where the communication structure among agents is modeled by a connected undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. Each vertex $i \in \mathcal{V}$ corresponds to an individual agent, and the edge

set $\mathcal{E} \subseteq \{(i, j) \mid i \neq j, i, j \in \mathcal{V}\}$ encodes the bidirectional communication links between agents. An edge $(i, j) \in \mathcal{E}$ is associated with a positive scalar weight $p_{ij} = p_{ji} > 0$, representing the interaction strength between agents i and j . The communication topology is further characterized by the adjacency matrix $P = [p_{ij}] \in \mathbb{R}^{n \times n}$, where $p_{ij} > 0$ if and only if $(i, j) \in \mathcal{E}$, and $p_{ij} = 0$ otherwise. For each agent $i \in \mathcal{V}$, we define its local neighborhood as $\mathcal{N}_i = \{j \in \mathcal{V} \mid (i, j) \in \mathcal{E}\}$. Communication is assumed to be local, i.e., each agent can exchange information only with its immediate neighbors.

3 Algorithm Development

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

$$\begin{aligned} \max_{\delta \in \mathbb{R}^m} \inf_{x \in X} L(x, \delta) &= \max_{\delta \in \mathbb{R}^m} \inf_{x \in X} \left\{ \sum_{i=1}^n (f_i(x_i) + \langle \delta, A_i x_i - d_i \rangle) \right\} \\ &= \max_{\delta \in \mathbb{R}^m} \sum_{i=1}^n \inf_{x_i \in X_i} \{f_i(x_i) + \langle \delta, A_i x_i - d_i \rangle\} \\ &= \max_{\delta \in \mathbb{R}^m} \sum_{i=1}^n -f_i^*(-\delta) - \langle \delta, d_i \rangle, \end{aligned} \quad (2)$$

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

$$f_i^*(\delta) = - \inf_{x_i \in X_i} \{f_i(x_i) - \langle \delta, A_i x_i \rangle\}.$$

Let $g_i(\delta) = f_i^*(-\delta) + \langle \delta, d_i \rangle$. Then, the dual problem in (2) can be written as $\min_{\delta \in \mathbb{R}^p} \sum_{i=1}^n g_i(\delta)$. To solve this problem in a distributed manner, each agent $i \in \mathcal{V}$ maintains a local copy y_i of the global variable δ . The resulting formulation is a distributed consensus optimization problem

$$\min_{y \in \mathbb{R}^{nm}} G(y) = \sum_{i=1}^n g_i(y_i), \quad \text{s. t. } y_i = y_j, \quad i, j \in \mathcal{V}, \quad (3)$$

where $y = [y_1^T \ y_2^T \ \dots \ y_n^T]^T$. In [25], it has been shown that, when the topology of the network is connected, problem (3) is equivalent to the following compact form:

$$\min_{y \in \mathbb{R}^{nm}} G(y) = \sum_{i=1}^n g_i(y_i), \quad \text{s. t. } W^{\frac{1}{2}} y = 0_{nm}. \quad (4)$$

where $W = \mathcal{L} \otimes I_m \succeq O_{nm}$ and \mathcal{L} is the Laplacian matrix of the network. By introducing the augmented Lagrangian function $L_\rho(y, v) = G(y) - v^T W^{\frac{1}{2}} y + \frac{\rho}{2} \|W^{\frac{1}{2}} y\|^2$, $\rho > 0$, we can linearize the method of multipliers [26] for solving (4), and the resulting update is

given as follows,

$$\begin{aligned} y^{k+1} &= \underset{y}{\operatorname{argmin}} \left\{ \langle \nabla G(y^k), y \rangle - \langle W^{\frac{1}{2}} v^k, y \rangle + \rho \langle W y^k, y \rangle \right. \\ &\quad \left. + \frac{\eta}{2} \|y - y^k\|^2 \right\} \\ &= y^k - \frac{1}{\eta} (\nabla G(y^k) + W^{\frac{1}{2}} v^k - \rho W y^k), \end{aligned} \quad (5a)$$

$$v^{k+1} = v^k - \rho W^{\frac{1}{2}} y^{k+1}, \quad (5b)$$

where $\nabla G(y^k) = [\nabla g_1^T(y_1^k) \nabla g_2^T(y_2^k) \dots \nabla g_n^T(y_n^k)]^T \in \mathbb{R}^{nm}$ and

$$\begin{aligned} \nabla g_i(y_i^k) &= -\nabla f_i^*(-y_i^k) + d_i \\ &= -\underset{z \in X_i}{\operatorname{argmin}} \{f_i(z) + \langle A_i z, y_i^k \rangle\} + d_i \\ &= -A_i x_i^{k+1} + d_i. \end{aligned} \quad (6)$$

Let $A \in \mathbb{R}^{nm \times np}$ be a block-diagonal matrix defined as $A = \operatorname{diag}(A_1, A_2, \dots, A_n)$ and $d = [d_1^T \ d_2^T \ \dots \ d_n^T]^T$. Then, $\nabla G(y^k) = Ax^k + d$, and the linearized method of multipliers for (4) is given as

$$x^{k+1} = \underset{z \in X}{\operatorname{argmin}} \{f(z) + \langle Az, y^k \rangle\}, \quad (7a)$$

$$y^{k+1} = y^k - \frac{1}{\eta} (-Ax^{k+1} + d - \lambda^k + \rho W y^k), \quad (7b)$$

$$\lambda^{k+1} = \lambda^k - \rho W y^{k+1}, \quad (7c)$$

where $\lambda^k = W^{\frac{1}{2}} v^k$. In the above algorithm, due to the presence of the local optimization problem at each iteration in (7a) which might not be efficiently solved, we approximate the solution of argmin in (7a) with the projection of $(x^k - \alpha(\nabla f(x^k) + A^T y^k))$ onto set X . Substituting the projection into the dual ascent steps yields our proposed algorithm:

$$x^{k+1} = \mathcal{P}_X(x^k - \alpha(\nabla f(x^k) + A^T y^k)), \quad (8a)$$

$$y^{k+1} = y^k - \frac{1}{\eta} (-Ax^{k+1} + d - \lambda^k + \rho W y^k), \quad (8b)$$

$$\lambda^{k+1} = \lambda^k - \rho W y^{k+1}, \quad (8c)$$

where \mathcal{P}_X is the projection operator over the convex set X . The distributed implementation of (8) over the undirected network $(\mathcal{V}, \mathcal{E})$ is detailed in Algorithm 1.

4 Convergence Analysis

In this section, we consider two cases, with and without local constraints, i.e., $X_i \subset \mathbb{R}^p$ and $X_i = \mathbb{R}^p$, $i \in \mathcal{V}$. For these two cases, we establish sublinear and linear convergence rates, respectively.

We denote the optimal solution of problem (1) by $h^* = [(x^*)^T \ (y^*)^T \ (\lambda^*)^T]^T$, where $x^* = [(x_1^*)^T \ (x_2^*)^T \ \dots]$

Algorithm 1 Distributed Gradient-based Algorithm

- 1: **Initialization:** $x^0 = [(x_1^0)^T \ (x_2^0)^T \ \dots \ (x_n^0)^T]^T$, $y^0 = [(y_1^0)^T \ (y_2^0)^T \ \dots \ (y_n^0)^T]^T$ are arbitrarily set. $\lambda^0 = [(\lambda_1^0)^T \ (\lambda_2^0)^T \ \dots \ (\lambda_n^0)^T]^T$ is initialized to satisfy $(1_m^T \otimes I_n) \lambda^0 = 0_m$. 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: For each agent $i \in \mathcal{V}$ computes $x_i^{k+1} = \mathcal{P}_{X_i}(x_i^k - \alpha(\nabla f(x_i^k) + A_i^T y_i^k))$.
 - 4: Each agent $i \in \mathcal{V}$ updates $y_i^{k+1} = y_i^k - \frac{1}{\eta} (-A_i x_i^{k+1} + d_i - \lambda_i^k + \rho t_{i,k})$ and sends the variable y_i^k to its neighbors $j \in \mathcal{N}_i$.
 - 5: 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 $t_i^{k+1} = \sum_{j \in \mathcal{N}_i} p_{ij} (y_i^{k+1} - y_j^{k+1})$.
 - 6: Each agent $i \in \mathcal{V}$ updates $\lambda_i^{k+1} = \lambda_i^k - t_i^{k+1}$.
 - 7: Set $k \leftarrow k + 1$ and go to Step 3 until a certain stopping criterion is satisfied, e.g., maximum number of iterations.
 - 8: **end for**
-

$(x_n^*)^T$, $y^* = [(y_1^*)^T \ (y_2^*)^T \ \dots \ (y_n^*)^T]^T$, and $\lambda^* = [(\lambda_1^*)^T \ (\lambda_2^*)^T \ \dots \ (\lambda_n^*)^T]^T$. The optimal solution satisfies the first-order optimality conditions (KKT conditions) for convex constrained optimization problems as described in [27]. In particular, the following conditions must hold. First, the dual variable y^* lies in the null space of the W , i.e.,

$$W y^* = 0_{nm}, \quad (9)$$

which indicates consensus of y_i^* for all $i \in \mathcal{V}$. Second, let $\sigma_i(x_i)$ be the indicator function associated with the convex set X_i , $i \in \mathcal{V}$, i.e., $\sigma_i(x_i) = 0$ if $x_i \in X_i$ and $\sigma_i(x_i) = +\infty$, otherwise. Then, the subdifferential condition associated with the objective and constraint functions is given by

$$-A^T y^* - \nabla f(x^*) \in \partial \sigma(x^*), \quad (10)$$

where $\sigma(x) = \sum_{i=1}^n \sigma_i(x_i)$ and $\partial \sigma(x^*)$ denotes the subdifferential of the indicator function encoding local constraints at x^* . This reflects a standard KKT-type optimality condition under nonsmooth composite objectives. In addition, based on the definition of $L_\rho(y, v)$ and (6), every v^* satisfying $W^{\frac{1}{2}} v^* = \nabla G(y^*) = -(Ax^* - d)$ is an optimal dual variable for problem (4), and hence,

$$\lambda^* = -(Ax^* - d). \quad (11)$$

Since λ^* is in the range space of W , (11) is equivalent to the coupled constraint, i.e., $(1_m^T \otimes I_n)(Ax^* - d) = 0_m$.

4.1 Sublinear Convergence Rate

In this section, we carry out the convergence analysis of the proposed algorithm under Assumptions 1 and 2. Denote the variable generated by the proposed algorithm at the k th iteration as $h^k = [(x^k)^T \ (y^k)^T \ (\lambda^k)^T]^T$. For $k \geq 0$, denote the difference of variables between the $(k+1)$ th iteration and the k th iteration as $\Delta x^{k+1} = x^{k+1} - x^k$, $\Delta y^{k+1} = y^{k+1} - y^k$, $\Delta \lambda^{k+1} = \lambda^{k+1} - \lambda^k$, $\Delta \nabla f(x^{k+1}) = \nabla f(x^{k+1}) - \nabla f(x^k)$ and $\Delta h^{k+1} = h^{k+1} - h^k$. The next two lemmas show the upper bound of the summation of $\|\Delta h^k\|_\Omega^2$ and the monotonicity of $\|\Delta h^k\|_\Omega^2$. The relationship between two consecutive iterations are given in Lemma 1.

Lemma 1 *Let Assumptions 1 and 2 hold. If $\frac{\rho}{\eta} < \frac{1}{\lambda_{\max}(W)}$ and $\alpha < \frac{1}{l_f}$, we have the following inequalities which establish the monotonicity of $\|h^k - h^*\|_\Omega^2$ and provide an upper bound of the summation of the sequence Δh^k generated by the proposed DGA, respectively,*

$$\begin{aligned} & \|h^k - h^*\|_\Omega^2 - \|h^{k+1} - h^*\|_\Omega^2 \\ & \geq \|h^{k+1} - h^k\|_\Omega^2 - \alpha l_f \|x^{k+1} - x^k\|^2, \quad k \geq 0, \end{aligned} \quad (12)$$

$$\sum_{t=0}^k \|\Delta h^{t+1}\|_\Omega^2 < \frac{\|h^0 - h^*\|_\Omega^2}{(1 - \alpha l_f)}, \quad k \geq 1, \quad (13)$$

where $\Omega \in \mathbb{R}^{(np+2nm) \times (np+2nm)}$ is defined as

$$\Omega = \begin{bmatrix} I_{np} & -\alpha A^T & O_{np \times nm} \\ O_{nm \times np} & \alpha \eta (I_{nm} - \frac{\rho}{\eta} W) & O_{nm \times nm} \\ O_{nm \times np} & O_{nm \times nm} & \frac{\alpha}{\rho} W^\dagger \end{bmatrix}.$$

PROOF. By (8b) and (9), we have

$$\lambda^{k+1} - \lambda^* = \lambda^k - \lambda^* - \rho W (y^{k+1} - y^*). \quad (14)$$

In addition, (8c) and (11) yield

$$\begin{aligned} y^{k+1} - y^* &= \left(I_{nm} - \frac{\rho}{\eta} W \right) (y^k - y^*) \\ &\quad - \frac{1}{\eta} (A(x^{k+1} - x^*) - (\lambda^k - \lambda^*)). \end{aligned} \quad (15)$$

Note that $(I_{nm} - \frac{\rho}{\eta} W) \succ O_{nm}$. Combining the above two equations leads to

$$\begin{aligned} & A(x^{k+1} - x^*) \\ &= \eta \left(I_{nm} - \frac{\rho}{\eta} W \right) (y^{k+1} - y^k) - (\lambda^{k+1} - \lambda^*). \end{aligned} \quad (16)$$

In addition, by the optimal condition of (8a), we have

$$-x^{k+1} + (x^k - \alpha(\nabla f(x^k) + A^T y^k)) \in \partial \sigma(x^{k+1}). \quad (17)$$

The convexity of the indicator function σ leads to $\langle \partial \sigma(x^{k+1}) - \partial \sigma(x^*), x^{k+1} - x^* \rangle \geq 0$. Therefore, in view of (10) and (17), we have

$$\begin{aligned} 0 &\leq \langle x^k - x^{k+1} - \alpha((\nabla f(x^k) + A^T y^k) - (A^T y^* + \nabla f(x^*))), x^{k+1} - x^* \rangle. \end{aligned} \quad (18)$$

Then, we can embed (16) into (18) to obtain

$$\begin{aligned} & -\alpha \langle \nabla f(x^k) - \nabla f(x^*), x^{k+1} - x^* \rangle \\ & \geq \langle x^{k+1} - x^k, x^{k+1} - x^* \rangle + \alpha \langle A^T(y^k - y^*), x^{k+1} - x^* \rangle \\ & = \langle x^{k+1} - x^k, x^{k+1} - x^* \rangle + \alpha \langle A^T(y^k - y^{k+1}), x^{k+1} - x^* \rangle \\ & \quad + \alpha \langle A^T(y^{k+1} - y^*), x^{k+1} - x^* \rangle \\ & = \langle x^{k+1} - x^k, x^{k+1} - x^* \rangle + \alpha \langle A^T(y^k - y^{k+1}), x^{k+1} - x^* \rangle \\ & \quad + \alpha \left\langle y^{k+1} - y^*, \eta \left(I_{nm} - \frac{\rho}{\eta} W \right) (y^{k+1} - y^k) \right\rangle \\ & \quad - \alpha \langle y^{k+1} - y^*, \lambda^{k+1} - \lambda^* \rangle. \end{aligned} \quad (19)$$

By (8c) and (9), we have $\lambda^{k+1} - \lambda^* = -\rho W (y^{k+1} - y^*)$. Since W is a semi-definite matrix with only one zero eigenvalue, we have the pseudo inverse matrix of W is also semi-definite with only one zero eigenvalue. Thus, we have

$$-\frac{1}{\rho} W^\dagger (\lambda^{k+1} - \lambda^k) = y^{k+1} - y^*, \quad (20)$$

and hence, $\langle y^* - y^{k+1}, \lambda^{k+1} - \lambda^* \rangle = \langle \frac{1}{\rho} W^\dagger (\lambda^{k+1} - \lambda^k), \lambda^{k+1} - \lambda^* \rangle$, substitution of which into (19) yields

$$\begin{aligned} & -\alpha \langle \nabla f(x^k) - \nabla f(x^*), x^{k+1} - x^* \rangle \\ & \geq \langle x^{k+1} - x^k, x^{k+1} - x^* \rangle + \alpha \langle A^T(y^k - y^{k+1}), x^{k+1} - x^* \rangle \\ & \quad + \alpha \left\langle y^{k+1} - y^*, \eta \left(I_{nm} - \frac{\rho}{\eta} W \right) (y^{k+1} - y^k) \right\rangle \\ & \quad + \frac{\alpha}{\rho} \langle W^\dagger (\lambda^{k+1} - \lambda^k), \lambda^{k+1} - \lambda^* \rangle \\ & = \langle h^{k+1} - h^k, h^{k+1} - h^* \rangle_\Omega \\ & = \frac{1}{2} (\|h^{k+1} - h^*\|_\Omega^2 - \|h^k - h^*\|_\Omega^2 + \|h^{k+1} - h^k\|_\Omega^2). \end{aligned} \quad (21)$$

By the Lipschitz smoothness and convexity of f , we have

$$f(x^{k+1}) \leq f(x^k) + \langle \nabla f(x^k), x^{k+1} - x^k \rangle + \frac{l_f}{2} \|x^{k+1} - x^k\|^2,$$

$$f(x^{k+1}) \geq f(x^*) + \langle \nabla f(x^*), x^{k+1} - x^* \rangle.$$

Combining the above two inequalities, we obtain

$$f(x^k) - f(x^*) + \frac{l_f}{2} \|x^{k+1} - x^k\|^2 \geq \langle \nabla f(x^*), x^{k+1} - x^* \rangle - \langle \nabla f(x^k), x^{k+1} - x^k \rangle,$$

which together with $f(x^k) - f(x^*) \leq \langle \nabla f(x^k), x^k - x^* \rangle$, results in

$$\langle \nabla f(x_k) - \nabla f(x^*), x_{k+1} - x^* \rangle \geq -\frac{l_f}{2} \|x^{k+1} - x^k\|^2,$$

substitution of which into (21) results in (12).

Noting that $\alpha < \frac{1}{l_f}$, we have

$$\|\Delta h^{k+1}\|_{\Omega - \alpha l_f \Omega'}^2 \geq (1 - \alpha l_f) \|\Delta h^{k+1}\|_\Omega^2 \geq 0, \quad (22)$$

where $\Omega' = \text{diag}\{I_{np}, O_{nm}, O_{nm}\}$ and the second equality holds at $\Delta h^{k+1} = 0$. Based on (22), summing (12) from 0 to k leads to (13). \square

Lemma 2 Let Assumptions 1 and 2 hold. If $\frac{\rho}{\eta} < \frac{1}{\lambda_{\max}(W)}$ and $\alpha < \frac{1}{l_f}$, then the sequence $\|\Delta h^k\|_\Omega$ generated by the proposed DGA algorithm is monotonically nonincreasing, that is,

$$\|\Delta h^k\|_\Omega^2 \geq \|\Delta h^{k+1}\|_\Omega^2. \quad (23)$$

PROOF. Similar to (16), we have the relationship among Δy^{k+1} , Δx^{k+1} and $\Delta \lambda^{k+1}$ as follows

$$\eta \left(I_{nm} - \frac{\rho}{\eta} W \right) (\Delta y^{k+1} - \Delta y^k) = A \Delta x^{k+1} - \Delta \lambda^{k+1}. \quad (24)$$

Taking the difference of the k th update and the $(k+1)$ th update of (17) yields

$$-(\Delta x^{k+1} - \Delta x^k) - \alpha(\Delta \nabla f(x^k) + A^T \Delta y^k) \in \sigma_{k+1} - \sigma_k,$$

which, in view of the convexity of σ , i.e., $\langle \Delta x^{k+1}, \sigma_{k+1} - \sigma_k \rangle \geq 0$, results in

$$\begin{aligned} & \langle \Delta x^{k+1}, -\Delta x^{k+1} + \Delta x^k \rangle + \alpha \langle \Delta x^{k+1}, -\Delta \nabla f(x^k) \rangle \\ & + \alpha \langle \Delta x^{k+1}, -A^T \Delta y^k \rangle \geq 0. \end{aligned} \quad (25)$$

The equivalent expression of the last term on the right hand side of the above equation is

$$\begin{aligned} & \alpha \langle \Delta x^{k+1}, -A^T \Delta y^k \rangle \\ & = \alpha \langle \Delta x^{k+1}, A^T(\Delta y^{k+1} - \Delta y^k) \rangle - \alpha \langle \Delta x^{k+1}, -A^T \Delta y^{k+1} \rangle \\ & = \alpha \langle \Delta x^{k+1}, A^T(\Delta y^{k+1} - \Delta y^k) \rangle \end{aligned}$$

$$\begin{aligned} & - \alpha \left\langle \eta \left(I_{nm} - \frac{\rho}{\eta} W \right) (\Delta y^{k+1} - \Delta y^k), -\Delta y^{k+1} \right\rangle \\ & - \alpha \langle \Delta \lambda^{k+1}, -\Delta y^{k+1} \rangle \\ & = \alpha \langle \Delta x^{k+1}, A^T(\Delta y^{k+1} - \Delta y^k) \rangle \\ & - \alpha \left\langle \eta \left(I_{nm} - \frac{\rho}{\eta} W \right) (\Delta y^{k+1} - \Delta y^k), -\Delta y^{k+1} \right\rangle \\ & - \alpha \left\langle \Delta \lambda^{k+1}, \frac{1}{\rho} W^\dagger (\Delta \lambda^{k+1} - \Delta \lambda^k) \right\rangle, \end{aligned} \quad (26)$$

where the second equality results from (24) and the last equality is due to (8c). Substituting (26) back into (25), we have

$$\begin{aligned} & -\alpha \langle \Delta \nabla f(x^k), \Delta x^{k+1} \rangle \\ & \geq \langle \Delta x^{k+1} - \Delta x^k, \Delta x^{k+1} \rangle + \alpha \langle A^T(\Delta y^k - \Delta y^{k+1}), \Delta x^{k+1} \rangle \\ & + \alpha \left\langle \eta \left(I_{nm} - \frac{\rho}{\eta} W \right) (\Delta y^{k+1} - \Delta y^k), \Delta y^{k+1} \right\rangle \\ & + \alpha \langle \Delta \lambda^{k+1} - \Delta \lambda^k, \Delta \lambda^{k+1} \rangle \\ & = \frac{1}{2} (\|\Delta h^{k+1}\|_\Omega^2 - \|\Delta h^k\|_\Omega^2 + \|\Delta h^{k+1} - \Delta h^k\|_\Omega^2). \end{aligned} \quad (27)$$

On the other hand

$$\begin{aligned} & -\alpha \langle \Delta \nabla f(x^k), \Delta x^{k+1} \rangle \\ & = -\alpha \langle \Delta \nabla f(x^k), \Delta x^{k+1} - \Delta x^k \rangle - \alpha \langle \Delta \nabla f(x^k), \Delta x^k \rangle \\ & \leq \left(\frac{\alpha}{2\xi} - \frac{\alpha}{l_f} \right) \|\Delta \nabla f(x^k)\|^2 + \frac{\alpha\xi}{2} \|\Delta x^{k+1} - \Delta x^k\|^2, \end{aligned} \quad (28)$$

where for any $\xi > 0$ the last inequality is due to the triangular inequality and the Lipschitz smoothness of f . Taking $\xi > l_f/2$ and combining (28) and (27) results in

$$\begin{aligned} & \|\Delta h^k\|_\Omega^2 \\ & \geq \|\Delta h^{k+1}\|_\Omega^2 + \|\Delta h^{k+1} - \Delta h^k\|_\Omega^2 - \alpha \xi \|\Delta x^{k+1} - \Delta x^k\|^2, \end{aligned}$$

which, in view of $\xi \in (\frac{l_f}{2}, \frac{1}{\alpha})$, completes the proof of Lemma 2. \square

Theorem 3 Let Assumptions 1 and 2 hold. If $\frac{\rho}{\eta} < \frac{1}{\lambda_{\max}(W)}$ and $\alpha \in \left(0, \min\left\{\frac{1}{l_f}, \frac{4\lambda_{\min}(\eta I_{nm} - \rho W)}{\|A\|_2^2}\right\}\right)$, then $\|\Delta h^k\|_\Omega^2 = o(1/k)$, and, consequently, the errors in the KKT conditions vanish at the rate of $o(1/k)$,

$$\left\| \sum_{i=1}^n A_i x_i^{k+1} - d_i \right\|^2 = o\left(\frac{1}{k}\right), \quad (29a)$$

$$\|\alpha(\nabla f(x^{k+1}) + A^T y^k) + \nu^{k+1}\|^2 = o\left(\frac{1}{k}\right), \quad (29b)$$

where $\nu^{k+1} \in \partial\sigma(x^{k+1})$.

PROOF. Combining Lemmas 1 and 2 with [16, Proposition 1], we have that the sequence Δh^k satisfies the non-ergodic rate, i.e., $\|\Delta h^k\|_\Omega^2 = o(1/k)$. Let $S = \frac{1}{2}(\Omega + \Omega^\top)$ and $\|\Delta h^k\|_\Omega^2 = \|\Delta h^k\|_S^2$. Since $\alpha < \frac{4\lambda_{\min}(\eta I_{nm} - \rho W)}{\|A\|_2^2}$, we have $\alpha\eta(I_{nm} - \frac{\rho}{\eta}W) - \frac{\alpha^2}{4}AA^\top \succeq 0$ and S is thus positive semidefinite, which directly implies that

$$\|x^{k+1} - x^k\|^2 = o\left(\frac{1}{k}\right), \quad (30a)$$

$$\|y^{k+1} - y^k\|_{\eta(I_{nm} - \frac{\rho}{\eta}W)}^2 = o\left(\frac{1}{k}\right). \quad (30b)$$

Since both λ^k and λ^* lie in the null space of W , left-multiplying both sides of (16) by $(1_m^\top \otimes I_n)$ gives

$$\begin{aligned} & \eta(1_m^\top \otimes I_n) \left(I_{nm} - \frac{\rho}{\eta}W \right) (y^{k+1} - y^k) \\ &= (1_m^\top \otimes I_n)(Ax^{k+1} - d) = \sum_{i=1}^n A_i x_i - d_i. \end{aligned}$$

Taking norms of both sides and using (30b), we obtain

$$\begin{aligned} & \left\| \sum_{i=1}^n A_i x_i^{k+1} - d_i \right\|^2 \\ &= \eta \|1_m^\top \otimes I_n\|^2 \|(I_{nm} - (\rho/\eta)W)\| \|y^{k+1} - y^k\|_{\eta(I_{nm} - \frac{\rho}{\eta}W)}^2, \end{aligned}$$

which establishes (29a). By (17), there exists $\nu^{k+1} \in \partial\sigma(x^{k+1})$ such that

$$\|x^{k+1} - x^k\| = \|\alpha(\nabla f(x^k) + A^\top y^k) + \nu^{k+1}\|,$$

Using the Lipschitz continuity of ∇f and inequality (30a) yields

$$\begin{aligned} & \|\alpha(\nabla f(x^{k+1}) + A^\top y^k) + \nu^{k+1}\| \\ &\leq \|\alpha(\nabla f(x^k) + A^\top y^k) + \nu^{k+1}\| + \|\nabla f(x^{k+1}) - \nabla f(x^k)\| \\ &\leq (\alpha l_f + 1) \|x^{k+1} - x^k\|, \end{aligned}$$

which establishes (29b). \square

4.2 Linear Convergence Rate

In this section, we consider the case of $X = \mathbb{R}^{np}$ and establish a linear convergence rate of Algorithm 1 under certain conditions.

Assumption 3 For each $i \in \mathcal{V}$, the cost function f_i is μ -strongly convex, i.e.,

$$\|\nabla f_i(x'_i) - \nabla f_i(x_i)\| \geq \mu \|x'_i - x_i\| \quad \forall x'_i, x_i \in \mathbb{R}^p,$$

for some $\mu > 0$.

Since the constraint in problem (1) is affine and X is convex, the feasible set is convex. Under a μ -strongly convex objective function and Assumption 2, the problem has a unique optimal solution in the feasible convex set.

Theorem 4 Suppose that $X = \mathbb{R}^{np}$ and Assumptions 1-3 hold. Suppose that the parameters satisfy

$$\eta > \max \left\{ \frac{\alpha \|A\|^2}{\left(\frac{2}{l_f + \mu} - \frac{1}{\theta_0} \right) \theta_2}, \frac{l_f + \mu}{2\mu l_f \alpha} \left(\rho \|A\|^2 \|W\| + \frac{\|A\|^2}{\theta_1} \right) \right\},$$

$\alpha < \frac{2}{l_f + \mu}$, $\theta_0 \in (\frac{l_f + \mu}{2}, \frac{1}{\alpha}]$, and $\theta_1 + \theta_2 < 1$, the sequence $\{h^k\}_{k \geq 0}$ generated by the proposed algorithm will converge linearly to the optimal solution h^* of the problem (1). In particular, we have

$$\|h^{k+1} - h^*\|_\Theta^2 \leq \delta \|h^k - h^*\|_\Theta^2, \quad (31)$$

where for some positive definite Θ and $\delta \in (0, 1)$.

PROOF. We first establish the basic inequality showing the decreasing property of the optimality gap $\|h^k - h^*\|_\Omega^2$ based on the smoothness and strong convexity of f . By (21), we have

$$\begin{aligned} & \|h^{k+1} - h^*\|_\Omega^2 - \|h^k - h^*\|_\Omega^2 \\ &\leq -\|\Delta h^{k+1}\|_\Omega^2 - 2\alpha \langle \nabla f(x^k) - \nabla f(x^*), x^{k+1} - x^* \rangle \\ &\leq -\|\Delta h^{k+1}\|_\Omega^2 - 2\alpha \langle \nabla f(x^k) - \nabla f(x^*), x^k - x^* \rangle \\ &\quad - 2\alpha \langle \nabla f(x^k) - \nabla f(x^*), \Delta x^{k+1} \rangle. \end{aligned} \quad (32)$$

For the last term on the right-hand side of the second inequality in (32), using Young's inequality, for any given $\theta_0 > 0$ we have

$$\begin{aligned} & \langle \nabla f(x^k) - \nabla f(x^*), \Delta x^{k+1} \rangle \\ &\geq -\frac{\|\nabla f(x^k) - \nabla f(x^*)\|^2}{2\theta_0} - \frac{\theta_0 \|\Delta x^{k+1}\|^2}{2}. \end{aligned} \quad (33)$$

Since f is l_f -smooth and μ -strongly convex, for the second term of (32) we have

$$\begin{aligned} & \langle \nabla f(x^k) - \nabla f(x^*), x^k - x^* \rangle \\ &\geq \frac{1}{l_f + \mu} \|\nabla f(x^k) - \nabla f(x^*)\|^2 + \frac{\mu l_f}{l_f + \mu} \|x^k - x^*\|^2. \end{aligned} \quad (34)$$

Using (33) and (34) in (32), we have

$$\begin{aligned} & \|h^{k+1} - h^*\|_\Omega^2 - \|h^k - h^*\|_\Omega^2 \\ &\leq -\|\Delta h^{k+1}\|_\Omega^2 + \alpha \theta_0 \|\Delta x^{k+1}\|^2 - \frac{2\mu l_f \alpha}{l_f + \mu} \|x^k - x^*\|^2 \\ &\quad - \left(\frac{2\alpha}{l_f + \mu} - \frac{\alpha}{\theta_0} \right) \|\nabla f(x^k) - \nabla f(x^*)\|^2. \end{aligned} \quad (35)$$

Taking square of both sides of (14) yields

$$\begin{aligned}\|\lambda^{k+1} - \lambda^*\|^2 &= \|\lambda^k - \lambda^* - \rho W(y^{k+1} - y^*)\|^2 \\ &= \|\lambda^k - \lambda^*\|^2 + \rho^2 \|y^{k+1} - y^*\|_{W^2}^2 \\ &\quad - 2\rho \langle W(y^{k+1} - y^*), \lambda^k - \lambda^* \rangle.\end{aligned}\quad (36)$$

Left-multiplying both sides of (15) by $\rho^{\frac{1}{2}}W^{\frac{1}{2}}$ yields, we have

$$\begin{aligned}&\rho^{\frac{1}{2}}W^{\frac{1}{2}} \left[(y^{k+1} - y^*) - \frac{1}{\eta}(\lambda^k - \lambda^*) \right] \\ &= \rho^{\frac{1}{2}}W^{\frac{1}{2}} \left[\left(I_{nm} - \frac{\rho W}{\eta} \right) (y^k - y^*) + \frac{1}{\eta} A(x^{k+1} - x^*) \right].\end{aligned}$$

Then, taking square both sides of the above equation results in

$$\begin{aligned}&\rho \|y^{k+1} - y^*\|_W^2 - 2 \left\langle \rho W(y^{k+1} - y^*), \frac{1}{\eta}(\lambda^k - \lambda^*) \right\rangle \\ &\quad + \frac{\rho}{\eta^2} \|\lambda^k - \lambda^*\|_W^2 \\ &= \rho \|y^k - y^*\|_{W(I_{nm} - \frac{\rho W}{\eta})^2}^2 + \frac{\rho}{\eta} \|A(x^{k+1} - x^*)\|_W^2 \\ &\quad + 2 \left\langle \rho W \left(I_{nm} - \frac{\rho W}{\eta} \right) (y^k - y^*), \frac{1}{\eta} A(x^{k+1} - x^*) \right\rangle \\ &= \rho \|(y^k - y^*)\|_{W(I_{nm} - \frac{\rho W}{\eta})^2}^2 + \frac{\rho}{\eta} \|A(x^{k+1} - x^*)\|_W^2 \\ &\quad + \frac{2}{\eta} \left\langle \rho W \left(I_{nm} - \frac{\rho W}{\eta} \right) (y^k - y^*), A(x^k - x^*) \right\rangle \\ &\quad - \frac{2}{\eta} \left\langle \rho W \left(I_{nm} - \frac{\rho W}{\eta} \right) (y^k - y^*), \alpha A(\nabla f(x^k) - \nabla f(x^*)) \right\rangle \\ &\quad - \frac{2}{\eta} \left\langle \rho W \left(I_{nm} - \frac{\rho W}{\eta} \right) (y^k - y^*), \alpha A A^T (y^k - y^*) \right\rangle \\ &\leq \rho \|y^k - y^*\|_{W(I_{nm} - \frac{\rho W}{\eta})^2}^2 + \frac{\rho}{\eta} \|A(x^{k+1} - x^*)\|_W^2 \\ &\quad + \frac{1}{\eta} \|y^k - y^*\|_{(\theta_1 + \theta_2)\rho^2 W^2 (I_{nm} - \frac{\rho W}{\eta})^2 - \alpha \rho A A^T W (I_{nm} - \frac{\rho W}{\eta})}^2 \\ &\quad + \frac{\|A\|^2}{\eta \theta_1} \|x^k - x^*\|^2 + \frac{\alpha^2 \|A\|^2}{\eta \theta_2} \|\nabla f(x^k) - \nabla f(x^*)\|^2 \\ &\leq \|y^k - y^*\|_H^2 + \frac{\rho \|A\|^2 \|W\|}{\eta} \|x^{k+1} - x^*\|^2 + \frac{\|A\|^2}{\eta \theta_1} \|x^k - x^*\|^2 \\ &\quad + \frac{\alpha^2 \|A\|^2}{\eta \theta_2} \|\nabla f(x^k) - \nabla f(x^*)\|^2,\end{aligned}\quad (37)$$

where $\theta_1, \theta_2 > 0$ and $H = \rho W \left(I_{nm} - \frac{\rho W}{\eta} \right)^2 + \frac{1}{\eta} \left[(\theta_1 + \theta_2) \rho^2 W^2 \left(I_{nm} - \frac{\rho W}{\eta} \right)^2 - \alpha A A^T \rho W \left(I_{nm} - \frac{\rho W}{\eta} \right) \right]$, the second equality results from (8a), and the inequality is by the application of the triangular inequality.

Combining (36) and (37) leads to

$$\|y^{k+1} - y^*\|_{\rho W - \frac{\rho^2}{\eta} W^2}^2 + \frac{1}{\eta} \|\lambda^{k+1} - \lambda^*\|^2$$

$$\begin{aligned}&- \frac{\rho \|A\|^2 \|W\|}{\eta} \|x^{k+1} - x^*\|^2 \\ &\leq \|\lambda^k - \lambda^*\|_{\frac{1}{\eta} I_{nm} - \frac{\rho}{\eta^2} W}^2 + \|y^k - y^*\|_H^2 + \frac{\|A\|^2}{\eta \theta_1} \|x^k - x^*\|^2 \\ &\quad + \frac{\alpha^2 \|A\|^2}{\eta \theta_2} \|\nabla f(x^k) - \nabla f(x^*)\|^2.\end{aligned}\quad (38)$$

Combining (35) and (38) leads to

$$\begin{aligned}&\left(1 - \frac{\rho \|A\|^2 \|W\|}{\eta} \right) \|x^{k+1} - x^*\|^2 + \|\lambda^{k+1} - \lambda^*\|_{\frac{\alpha}{\rho} W^\dagger + \frac{1}{\eta} I_{nm}}^2 \\ &\quad + \|y^{k+1} - y^*\|_{(\alpha \eta I_{nm} + \rho W)(I_{nm} - \frac{\rho}{\eta} W)}^2 \\ &\leq \left(1 + \frac{\|A\|^2}{\eta \theta_1} - \frac{2 \mu l_f \alpha}{l_f + \mu} \right) \|x^k - x^*\|^2 \\ &\quad + \|y^k - y^*\|_{\alpha \eta (I_{nm} - \frac{\rho}{\eta} W) + H}^2 + \|\lambda^k - \lambda^*\|_{\frac{\alpha}{\rho} W^\dagger + \frac{1}{\eta} I_{nm} - \frac{\rho}{\eta^2} W}^2 \\ &\quad - \|\Delta h^{k+1}\|_\Omega^2 + \alpha \theta_0 \|\Delta x^{k+1}\|^2 \\ &\quad - \left(\frac{2 \alpha}{l_f + \mu} - \frac{\alpha}{\theta_0} - \frac{\alpha^2 \|A\|^2}{\eta \theta_2} \right) \|\nabla f(x^k) - \nabla f(x^*)\|^2.\end{aligned}\quad (39)$$

To prove that there exists a $\delta \in (0, 1)$ such that (31) holds, we first show that the following inequalities hold,

$$-\frac{\rho \|A\|^2 \|W\|}{\eta} \succ \frac{\|A\|^2}{\eta \theta_1} - \frac{2 \mu l_f \alpha}{l_f + \mu}, \quad (40a)$$

$$(\alpha \eta I_{nm} + \rho W) \left(I_{nm} - \frac{\rho}{\eta} W \right) \succ \alpha \eta \left(I_{nm} - \frac{\rho}{\eta} W \right) + H, \quad (40b)$$

$$\frac{\alpha}{\rho} W^\dagger + \frac{1}{\eta} I_{nm} \succ \frac{\alpha}{\rho} W^\dagger + \frac{1}{\eta} I_{nm} - \frac{\rho}{\eta^2} W, \quad (40c)$$

$$1 - \alpha \theta_0 \geq 0, \quad (40d)$$

$$\frac{2 \alpha}{l_f + \mu} - \frac{\alpha}{\theta_0} - \frac{\alpha^2 \|A\|^2}{\eta \theta_2} \geq 0. \quad (40e)$$

Inequality (40a) holds by the condition of the theorem, that is, $\eta > \frac{l_f + \mu}{2 \mu l_f \alpha} \left(\rho \|A\|^2 \|W\| + \frac{\|A\|^2}{\theta_1} \right)$. (40b) is equivalent to

$$\begin{aligned}&\frac{\rho}{\eta} W \left(I_{nm} - \frac{\rho W}{\eta} \right) \\ &\quad \left(\alpha A A^T + \rho W - (\theta_1 + \theta_2) \rho W \left(I_{nm} - \frac{\rho W}{\eta} \right) \right) \succ O_{nm \times nm},\end{aligned}$$

which is true if the following inequality holds

$$\alpha A A^T + \rho (1 - \theta_1 - \theta_2) W + \frac{\rho^2 (\theta_1 + \theta_2) W}{\eta} \succ O_{nm \times nm}.$$

Since W and $A A^T$ are positive semidefinite, $\theta_1 + \theta_2 < 1$ ensures the above inequality and hence (40b). Inequality (40c) is true due to the positiveness of both ρ and η . Setting $\alpha < \frac{2}{l_f + \mu}$, $\theta_0 \in (\frac{l_f + \mu}{2}, \frac{1}{\alpha}]$ and $\eta \geq \frac{\alpha \|A\|^2}{(\frac{l_f + \mu}{2} - \frac{1}{\theta_0}) \theta_2}$ ensures (40d) and (40e).

Based on inequalities (40d) and (40e), we can eliminate the nonpositive terms on the right-hand side of (39), leading to the following recursive inequality that relates the $(k+1)$ th and k th iterations,

$$\|h^{k+1} - h^*\|_\Theta^2 \leq \|h^k - h^*\|_\Theta^2, \quad (41)$$

where Θ and Θ' are block-diagonal matrices defined as $\Theta = \text{diag}\{\Theta_1, \Theta_2, \Theta_3\}$ and $\Theta' = \text{diag}\{\Theta_1, \Theta_2, \Theta_3\}$ with individual blocks given by $\Theta_1 = I_{nm} - \frac{\rho\|A\|^2\|W\|}{\eta}$, $\Theta_2 = (\alpha\eta I_{nm} + \rho W)(I_{nm} - \frac{\rho}{\eta}W)$, $\Theta_3 = \frac{\alpha}{\rho}W^\dagger + \frac{1}{\eta}I_{nm}$, $\Theta'_1 = (1 + \frac{\|A\|^2}{\eta\theta_1} - \frac{2\mu l_f \alpha}{l_f + \mu})I_{nm}$, $\Theta'_2 = \alpha\eta(I_{nm} - \frac{\rho}{\eta}W) + H$, and $\Theta'_3 = \frac{\alpha}{\rho}W^\dagger + \frac{1}{\eta}I - \frac{\rho}{\eta^2}W$. Furthermore, in view of the fact that inequalities (40a), (40b), and (40c) hold by proper parameter selections, we have $\Theta \succ \Theta' \succeq O_p$. Consequently, there exists a scalar $\delta \in (0, 1)$ such that $\delta\Theta \succeq \Theta'$. Substituting this bound into (41), we obtain the following inequality,

$$\begin{aligned} \|h^{k+1} - h^*\|_\Theta^2 &\leq \delta \|h^k - h^*\|_\Theta^2 + \|h^k - h^*\|_{\Theta' - \delta\Theta}^2 \\ &\leq \delta \|h^k - h^*\|_\Theta^2, \end{aligned}$$

which completes the proof. \square

5 Numerical Experiments

To validate the effectiveness of the proposed algorithm and to compare its performance with existing methods, we conduct a series of numerical experiments on a standard benchmark system. Specifically, the proposed method is compared with five state-of-the-art distributed algorithms: the Augmented Lagrangian Tracking (ALT) algorithm from [15], Mirror-PG-EXTRA and Mirror-P-EXTRA from [16], DMAC from [19], and DuSPA from [17]. Notably, the proposed algorithm and Mirror-PG-EXTRA rely solely on gradient information, while Mirror-P-EXTRA, DMAC, and ALT require solving local optimization subproblems via the argmin operator at each iteration.

We consider the economic dispatch problem over the IEEE 118-bus system [28], where the communication network is modeled as an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with $\mathcal{V} = 1, 2, \dots, 118$ and edges defined by $(i, i+1)$ and $(i, i+2)$ for $1 \leq i \leq 116$. This yields a sparse yet connected topology. A doubly stochastic weight matrix is generated accordingly. Among the 118 buses, 14 are randomly designated as generator buses, denoted by the set \mathcal{V}_g . As for the settings for the optimization, we adapt the emission cost function in [21] to model the following optimization problem, whose local cost of each generator is a combination of quadratic functions and exponential

functions

$$\begin{aligned} \min_P f(P) &= \sum_{i=1}^{118} a_i P_i^2 + b_i P_i + \delta_i e^{\ell_i P_i}, \quad (42) \\ \text{s. t. } \sum_{i=1}^{118} P_i &= d, \quad \underline{p}_i \leq P \leq \bar{p}_i, \quad \forall i \in \mathcal{V}, \end{aligned}$$

where $P = [P_1 \ P_2 \ \dots \ P_{118}]^T$ and the coefficients for each generator are randomly generated positive numbers. To be more specific, for $i \in \mathcal{V}_g$, $a_i \sim \mathcal{U}([0.3, 0.7]$, $b_i \sim \mathcal{U}([100, 400])$, $\delta_i \sim \mathcal{U}([1, 10]) \times 10^{-4}$ and $\ell_i \sim \mathcal{U}([1, 10]) \times 10^{-2}$. As for the limitations of generators, we set $[\underline{p}_i, \bar{p}_i] = [0, 250]$, $i \in \mathcal{V}_g$. For buses without generators, i.e., for $i \in \mathcal{V} \setminus \mathcal{V}_g$, their corresponding coefficients a_i , b_i and δ_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 the summation 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. The optimal solution P^* of this problem is obtained by solving the problem centrally using a state-of-the-art solver MOSEK in YALMIP [29], and is used as a reference for evaluating convergence and accuracy. For the argmin operation in Mirror-P-EXTRA, DMAC, and ALT, we also invoke the solver MOSEK in YALMIP to get the solutions of local optimization problems. Our experimental setup uses a Laptop with an Intel Core i7-13700HX CPU, and the MATLAB version is R2023b.

Figure 1 illustrates the evolution of the generator outputs, showing the difference between each P_i and its optimal value P_i^* over iterations. By the 300th iteration, all generators exhibit negligible error, indicating convergence to the optimal dispatch. Figure 2 presents the power mismatch $\|P_k - P^*\|^2$ across iterations for the five algorithms. The proposed method consistently outperforms all baselines in terms of both convergence speed and solution quality. To assess computational time efficiency, Figure 3 plots the optimality error over actual computation time. It is observed that the proposed algorithm achieves the fastest convergence, significantly outperforming Mirror-P-EXTRA, Mirror-PG-EXTRA, DMAC, DuSPA, and ALT. This performance advantage arises because these methods require solving local optimization problems at each iteration. In contrast, the proposed algorithm, Mirror-PG-EXTRA, and DuSPA only involve first-order gradient computations and simple projections, resulting in superior time efficiency and faster error decay.

The simulation results on the IEEE 118-bus system demonstrate the effectiveness and efficiency of the proposed algorithm. Compared to five representative state-of-the-art methods, our algorithm achieves superior

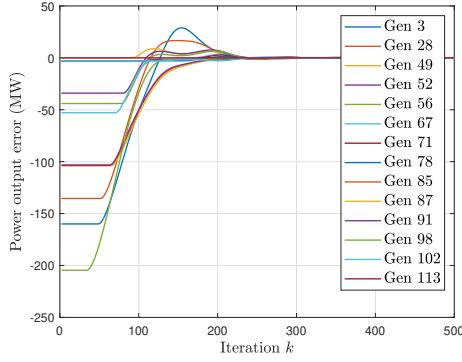


Fig. 1. The power output error of each generator under the proposed Algorithm 1.

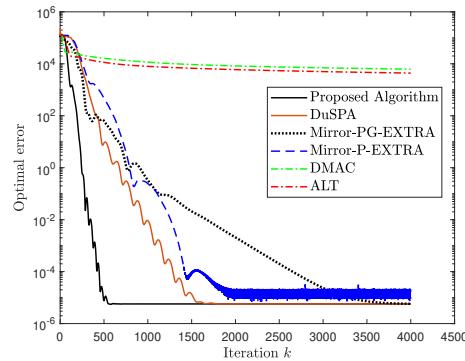


Fig. 2. Comparison of optimal error over iteration for the six algorithms.

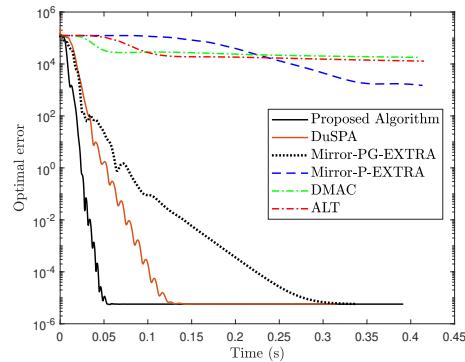


Fig. 3. Comparison of optimal error over computation time for the six algorithms.

performance in terms of convergence speed and solution accuracy, both with respect to iteration count and CPU time. This advantage is primarily attributed to the algorithm's lightweight per-iteration computation, which avoids solving local optimization subproblems and instead leverages gradient-based updates with simple projection steps. Furthermore, the proposed method maintains competitive accuracy while offering better scalability and practicality for resource-constrained

distributed energy systems.

6 Conclusions

This paper presents a lightweight and scalable distributed algorithm for solving optimization problems with coupling equality constraints over networked systems. Unlike many existing methods that require solving local optimization problems at every iteration, the proposed algorithm obviates such a need by employing a projection-based gradient update. As a result, the proposed algorithm forms a single-loop iteration and reduces per-iteration computational overhead while preserving convergence guarantees. We rigorously analyze the convergence behavior under both convex and strongly convex cost function settings, establishing nonergodic sublinear and linear convergence rates for general convex and strongly convex cases, respectively. Extensive numerical experiments on the IEEE 118-bus system validate the algorithm's effectiveness and demonstrate its superior performance in both convergence speed and computational efficiency compared to several benchmark methods.

References

- [1] Yijing Xie, Yichen Zhang, Wei-Jen Lee, Zongli Lin, and Yacov A Shamash. Virtual power plants for grid resilience: A concise overview of research and applications. *IEEE/CAA Journal of Automatica Sinica*, 11(2):329–343, 2024.
- [2] Haitao Zhao, Kaiqi Ding, Nurul I Sarkar, Jibo Wei, and Jun Xiong. A simple distributed channel allocation algorithm for D2D communication pairs. *IEEE Transactions on Vehicular Technology*, 67(11):10960–10969, 2018.
- [3] Bushra Jamil, Humaira Ijaz, Mohammad Shojafar, Kashif Munir, and Rajkumar Buyya. Resource allocation and task scheduling in fog computing and internet of everything environments: A taxonomy, review, and future directions. *ACM Computing Surveys (CSUR)*, 54(11s):1–38, 2022.
- [4] John Hetzer, C Yu David, and Kalu Bhattacharai. An economic dispatch model incorporating wind power. *IEEE Transactions on Energy Conversion*, 23(2):603–611, 2008.
- [5] Mudathir Funsho Akorede, Hashim Hizam, and Edris Poutesmaeil. Distributed energy resources and benefits to the environment. *Renewable and Sustainable Energy Reviews*, 14(2):724–734, 2010.
- [6] Abir Muhtadi, Dilip Pandit, Nga Nguyen, and Joydeep Mitra. Distributed energy resources based microgrid: Review of architecture, control, and reliability. *IEEE Transactions on Industry Applications*, 57(3):2223–2235, 2021.
- [7] Rui Wang, Qiqiang Li, Bingying Zhang, and Luhao Wang. Distributed consensus based algorithm for economic dispatch in a microgrid. *IEEE Transactions on Smart Grid*, 10(4):3630–3640, 2019.
- [8] Xiao-Kang Liu, Jiaqi Yan, Lantao Xing, and Changyun Wen. Constrained consensus-based iterative algorithm for economic dispatch in power systems. In *The 46th Annual Conference of the IEEE Industrial Electronics Society*, pages 3537–3542. IEEE, 2020.
- [9] Wei Chen and Guo-Ping Liu. Privacy-preserving consensus-based distributed economic dispatch of smart grids via state decomposition. *IEEE/CAA Journal of Automatica Sinica*, 11(5):1250–1261, 2024.

- [10] Lianghao Ji, Yuhe Dou, Cuijuan Zhang, and Huaqing Li. Self-triggered consensus-based strategy for economic dispatch in uncertain communication networks. *IEEE Transactions on Network Science and Engineering*, 2024.
- [11] Tsung-Hui Chang, Mingyi Hong, and Xiangfeng Wang. Multi-agent distributed optimization via inexact consensus ADMM. *IEEE Transactions on Signal Processing*, 63(2):482–497, 2014.
- [12] Gang Chen and Qing Yang. An ADMM-based distributed algorithm for economic dispatch in islanded microgrids. *IEEE Transactions on Industrial Informatics*, 14(9):3892–3903, 2017.
- [13] Lei Sun, Derui Ding, Hongli Dong, and Xiaojian Yi. Distributed economic dispatch of microgrids based on admm algorithms with encryption-decryption rules. *IEEE Transactions on Automation Science and Engineering*, 2024.
- [14] Yanxu Su, Qingling Wang, and Changyin Sun. Distributed primal-dual method for convex optimization with coupled constraints. *IEEE Transactions on Signal Processing*, 70:523–535, 2022.
- [15] Alessandro Falsone and Maria Prandini. Augmented lagrangian tracking for distributed optimization with equality and inequality coupling constraints. *Automatica*, 157:111269, 2023.
- [16] Angelia Nedić, Alex Olshevsky, and Wei Shi. Improved convergence rates for distributed resource allocation. In *Proceedings of the IEEE Conference on Decision and Control (CDC)*, pages 172–177, 2018.
- [17] Jinming Xu, Shanying Zhu, Yeng Chai Soh, and Lihua Xie. A dual splitting approach for distributed resource allocation with regularization. *IEEE Transactions on Control of Network Systems*, 6(1):403–414, 2019.
- [18] Jiaqi Zhang, Keyou You, and Kai Cai. Distributed dual gradient tracking for resource allocation in unbalanced networks. *IEEE Transactions on Signal Processing*, 68:2186–2198, 2020.
- [19] Wenwen Wu, Shanying Zhu, Shuai Liu, and Xinping Guan. Distributed mismatch tracking algorithm for constraint-coupled resource allocation: Optimality and differential privacy. *IEEE Transactions on Automatic Control*, 2025.
- [20] Huaqing Li, Zheng Wang, Guo Chen, and Zhao Yang Dong. Distributed robust algorithm for economic dispatch in smart grids over general unbalanced directed networks. *IEEE Transactions on Industrial Informatics*, 16(7):4322–4332, 2020.
- [21] Xing He, You Zhao, and Tingwen Huang. Optimizing the dynamic economic dispatch problem by the distributed consensus-based admm approach. *IEEE Transactions on Industrial Informatics*, 16(5):3210–3221, 2019.
- [22] Shanying Zhu, Tao Ding, Cailian Chen, Mo-Yuen Chow, and Xinping Guan. Deed-admm: A scalable distributed algorithm for economic dispatch in multi-energy systems with energy storage. *IEEE Transactions on Automation Science and Engineering*, 2025.
- [23] Tao Yang, Jie Lu, Di Wu, Junfeng Wu, Guodong Shi, Ziyang Meng, and Karl Henrik Johansson. A distributed algorithm for economic dispatch over time-varying directed networks with delays. *IEEE Transactions on Industrial Electronics*, 64(6):5095–5106, 2016.
- [24] Xiasheng Shi, Changyin Sun, and Chaoxu Mu. Distributed economic dispatch algorithm with quantized communication mechanism. *IEEE Transactions on Automation Science and Engineering*, 2024.
- [25] Chenyang Qiu, Shanying Zhu, Zichong Ou, and Jie Lu. A stochastic second-order proximal method for distributed optimization. *IEEE Control Systems Letters*, 7:1405–1410, 2023.
- [26] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, Jonathan Eckstein, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*, 3(1):1–122, 2011.
- [27] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.
- [28] P Venkatesh, R Gnanadass, and Narayana Prasad Padhy. Comparison and application of evolutionary programming techniques to combined economic emission dispatch with line flow constraints. *IEEE Transactions on Power Systems*, 18(2):688–697, 2003.
- [29] Johan Löfberg. Yalmip: A toolbox for modeling and optimization in matlab. In *In Proceedings of the CACSD Conference*, Taipei, Taiwan, 2004.