2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

Fast Distributed Resource Allocation of Smart Grid: A Zeroth-Order Optimization Algorithm

Meng Luan Guanghui Wen Wei Xing Zheng

Abstract— **This paper investigates a multi-objective distributed resource allocation problem, where the economic cost including the transmission loss, and the environmental pollution are taken into account simultaneously. To settle this problem, a Pareto-based zeroth-order fast distributed optimization algorithm is proposed, which can always balance the overall energy demand with generation. In the algorithm design, the acceleration idea of the momentum method is tailored for the gradient estimation update, which gives a more accurate descent direction. Moreover, the unknown effect causes the gradient of the objective function to be unavailable and only the function values to be observed. Different from the gradient-based methods, a zeroth-order method is proposed to solve the distributed resource allocation problem with gradient estimation. Furthermore, the convergence of the designed algorithm is proved theoretically, and the convergence rate of linear speedup can be achieved. Finally, numerical simulations verify the validity and applicability of the proposed algorithm.**

I. INTRODUCTION

Over the past few years, distributed resource allocation (DRA) [1] as an emerging research field has gained increasing attention to mitigate the greenhouse effect. DRA is a framework for coordinating the operation of multiple distributed energy resources to optimize energy delivery and consumption. It is regarded as a significant tool for the smart grid modernization, which allows the integration of renewable energy sources and improves the efficiency, reliability, and security of the power grid [2].

Many centralized algorithms have been designed to address the DRA problem. They can be mainly divided into the heuristic-based algorithms such as the particle swarm optimization [3], and the analysis-based algorithms including the linear programming [4]. Additionally, research has been conducted for DRA resorting to the artificial intelligence and machine learning approaches [5]. However, the aforementioned approaches are usually limited in scalability and inefficient. Specially, they can be easily fooled or misled by local optimums, resulting in suboptimal solutions. These limitations have prompted the shift from centralized control to distributed management in the power development industry.

This work was supported in part by the National Key Research and Development Program of China under Grant No. 2022YFA1004702, in part by the National Natural Science Foundation of China through Grant Nos. U22B2046, 62073079 and 62088101, and in part by the General Joint Fund of the Equipment Advance Research Program of Ministry of Education under Grant No. 8091B022114.

M. Luan and G. Wen are with the Department of Systems Science, School of Mathematics, Southeast University, Nanjing 211189, China (e-mail: 228433lm@seu.edu.cn; ghwen@seu.edu.cn).

W. X. Zheng is with the School of Computer, Data and Mathematical Sciences, Western Sydney University, Sydney, NSW 2751, Australia (email: w.zheng@westernsydney.edu.au).

In recent years, distributed optimization algorithms have been extensively studied to tackle the weakness of the centralized methods [6]. For the next generation power grid, one of the fundamental and important problems of smart grid is to balance the energy demand with generation, which involves handling the coupling constraints. To this end, the continuous-time dynamic primal-dual method was developed for optimization problems with affine equality constraints in [7], while the dual decomposition method was utilized in [8]. However, these algorithms are equipped with slower convergence rates, and the introduction of dual variables raises the requirements of computation and storage. To further accelerate the convergence speed, the EXTRA algorithm [9], the heavy ball method [10], the Nesterov gradient descent algorithm [11], and the gradient tracking method [12] were proposed. Despite these algorithms, it is still challenging to fast handle optimization problems with coupling constraints. This paper aims to fill this gap. Furthermore, these accelerated works mainly focused on the gradient-based algorithm research. Nowadays it is still an open problem on how to realize the fast response of DRA when the gradient information is unavailable.

To enable the power grid to operate in an efficient mode, the gradient information of the objective function is generally utilized to calculate the optimal operating points for generators. It can maximize the efficiency of the power system and reduce the risk of power outages and other system failures. In practice, smart grid integrates the renewable energy sources into the power system. These renewable energy sources are highly dependent on the weather, which leads to the incompleteness of data and the difficulties in obtaining the gradient information. Besides, the environmental pollution from the power generation can only be observed with the limited observation. On the other hand, it is noteworthy that there is a growing interest in the zeroth-order distributed optimization algorithms [13] when the gradient information is not available. These situations usually occur when the gradient of the objective function is infeasible or costly to evaluate due to the limited access of distributed devices, or when only the black-box programs are available to calculate the function values. Moreover, the zeroth-order distributed optimization algorithms require a small amount of computation and can fast achieve a high-quality solution [14], which is also suitable for large-scale resource allocation. Concretely, the complexity bounds and the convergence rate for the zeroth-order methods of convex optimization were analyzed in [15]. The zeroth-order idea was designed into the online alternating direction method of multipliers (ADMM)

[16]. Thus, it is very interesting to adopt the zeroth-order method to realize the fast demand response for DRA.

Motivated by the above analysis, a zeroth-order fast distributed optimization algorithm is proposed for DRA in this paper. The main contributions are illustrated in the following aspects. Primarily, a novel multi-objective DRA problem is introduced, which integrates the economic cost and the environmental pollution simultaneously. In addition, the transmission loss is taken into account in the economic cost. Next, for the considered optimization problem, a zeroth-order fast distributed optimization algorithm is developed. The momentum method is employed to accelerate the convergence rate, and the zeroth-order scheme is utilized to address the unknown gradient information. Moreover, it is theoretically proved that the designed algorithm can achieve a linear-speedup convergence rate. Compared with the previous works, this new algorithm accelerates the convergence on the premise of satisfying the coupling constraints.

The rest of this paper is organized as follows. Section II introduces the graph theory, useful lemmas, and the problem model. Section III presents the designed algorithm and the convergence analysis. Section IV carries out simulation experiments. Finally, Section V summarizes the whole paper.

Notation. $\mathbb{R}$, $\mathbb{R}^n$ and $\mathbb{R}^{m\times n}$ represent the spaces of real numbers, n-dimensional real vectors and $m\times n$ real matrices, respectively. $\|\cdot\|$ denotes the Euclidian norm. $\mathbf{1}$ denotes the column vector with all elements being 1. $\nabla f(x)$ and $\nabla^2 f(x)$ denote the gradient and Hessian matrix of function $f(x)$. For a real and symmetric matrix M, $M \preceq 0$ (or $M \succeq 0$) means that it is negative (or positive) semidefinite. For a differentiable function $f(\cdot)$ and for all $u, v \in \mathbb{R}^p$, it is l_1-strongly convex if $f(v) \geq f(u) + \langle \nabla f(u), v-u\rangle + \frac{l_1}{2}\|v-u\|^2$, and it is l_2-smooth if $\|\nabla f(u) - \nabla f(v)\| \leq l_2 \|u-v\|$.

II. PRELIMINARIES AND PROBLEM FORMULATION

A. Preliminaries

An undirected graph consisting of N nodes is depicted as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{1, 2, \cdots, N\}$ and $\mathcal{E} \subseteq \mathcal{V}\times\mathcal{V}$ denote the node set and the edge set, respectively. If nodes i and j can communicate with each other, then $(i,j) \in \mathcal{E}$. The adjacency matrix $W = [w_{ij}] \in \mathbb{R}^{N\times N}$ of graph $\mathcal{G}$ is defined as $w_{ij} = w_{ji} = 1$ if $(i,j) \in \mathcal{E}$, and $w_{ij} = 0$ otherwise. Specially, $w_{ii} = 0$ for $i = 1, 2, \cdots, N$. The Laplacian matrix L is defined as $L = D - W$, where $D \in \mathbb{R}^{N\times N}$ is a diagonal matrix with $D_{ii} = \sum_{j=1}^{N} w_{ij}$, $i \in \{1, 2, \cdots, N\}$ being diagonal elements. A path from node i_0 to node i_m is a sequence of nodes $\{i_0, i_1, \cdots, i_m\}$ satisfying $(i_j, i_{j+1}) \in \mathcal{E}$, $j = 0, 1, \cdots, m-1$. Graph $\mathcal{G}$ is called connected if each node in $\mathcal{G}$ has a path to any other node.

Assumption 1: Graph $\mathcal{G}$ is undirected and connected.

The lemma presented below is beneficial to the convergence analysis.

Lemma 1: Suppose that Q is a symmetric matrix with its eigenvalues $\lambda_1, \cdots, \lambda_N$ satisfying $\lambda_1 \leq \cdots \leq \lambda_N$ and the associated eigenvectors being $u_1, \cdots, u_N$. Then $x^T Q x \geq \lambda_i x^T x$ holds for all $x \in \{x \mid x \perp u_j,\ j = 1, 2, \cdots, i-1\}$, $i = 1, 2, \cdots, N$.

Then, the deterministic gradient estimator for a differentiable function $f(x) : \mathbb{R}^p \mapsto \mathbb{R}$ is designed as follows [17]:

$$\widehat{\nabla}_\delta f(x) = \frac{1}{\delta}\sum_{l=1}^{p}\left(f(x+\delta \mathbf{e}_l) - f(x)\right)\mathbf{e}_l, \tag{1}$$

where $\delta > 0$ is a parameter and $\mathbf{e}_l$ is a unit vector with the lth element being 1 and the others being 0. The gradient estimator $\widehat{\nabla}_\delta f(x)$ can be obtained through sampling the function values of $f(x)$ at $p+1$ points. The following lemma reveals that $\widehat{\nabla}_\delta f(x)$ can be very close to $\nabla f(x)$ for a sufficiently small δ.

Lemma 2 ([17]): For a l-smooth function $f(x)$, $\forall x \in \mathbb{R}^p$ and $\forall \delta > 0$, there holds

$$\left\|\widehat{\nabla}_\delta f(x) - \nabla f(x)\right\| \leq \frac{\sqrt{p}l\delta}{2}. \tag{2}$$

B. Problem Formulation

A novel DRA is proposed in this part, which aims at minimizing the economic cost and the environmental pollution simultaneously for the power system with N generators on the premise of satisfying the balance between the energy supply and the demand. The involved cost functions are modeled as follows.

1) *Economic cost*: The economic cost mainly consists of two parts: the generation cost for traditional generators or the operation cost for wind generators, and the transmission loss. Many geographically distributed controllable devices raise new challenges for DRA, where the transmission line loss has a distinct impact on the electricity scheduling due to the long-distance transmission. Thus, the economic cost for generator i, $i = 1, \cdots, N$, is modeled as follows [18], [19]:

$$f_i^{\mathrm{eco}}(P_i) = a_i P_i^2 + b_i P_i + c_i + d_i {P_i}^2, \tag{3}$$

where $a_i > 0$, b_i, and c_i are the constant coefficients, P_i is the supply of generator i, and d_i ($0 < d_i < 1$) is the constant coefficient of the transmission loss.

2) *Environmental pollution cost:* The future power system will be transformed to green and low carbon, which puts forward higher requirements for gas emissions. During the operation of conventional generators, the emission of harmful gases generated by generator i including SO_x, CO_x and others can be modeled as follows [20], [21]:

$$f_i^{\mathrm{env}}(P_i) = \tilde{a}_i P_i^2 + \tilde{b}_i P_i + \tilde{c}_i, \tag{4}$$

where $\tilde{a}_i > 0$, $\tilde{b}_i$, and $\tilde{c}_i$ are the constant coefficients.

According to the above analysis, combined with the supply and demand balance constraints, the multi-objective optimization problem is established as follows:

$$\min \left[\sum_{i=1}^{N} f_i^{\mathrm{eco}}(P_i),\ \sum_{i=1}^{N} f_i^{\mathrm{env}}(P_i)\right] \tag{5a}$$

$$s.t.\ \ \sum_{i=1}^{N} P_i = D, \tag{5b}$$

where constant D is the total demand. The contradiction between multiple costs indicates that it is impossible to minimize these goals simultaneously. Resorting to the Pareto optimality [22], we transform this problem into a single objective optimization problem through a weighted approach, and further abbreviated as

$$\min \quad f(\boldsymbol{x}) \triangleq \sum_{i=1}^{N} f_i(x_i) \tag{6a}$$

$$s.t. \quad \sum_{i=1}^{N} x_i = D, \tag{6b}$$

where the decision variable x_i is employed to represent P_i in (5) for convenience, $f_i(x_i) = \eta_1 f_i^{\mathrm{eco}}(x_i) + \eta_2 f_i^{\mathrm{env}}(x_i)$, and the coefficients $\eta_1, \eta_2 \in (0,1)$ satisfies $\eta_1 + \eta_2 = 1$. Clearly, $f(\boldsymbol{x})$ is strongly-convex and smooth, thus we can let $l_1 I \leq \nabla^2 f(\boldsymbol{x}) \leq l_2 I$, where l_1 and l_2 are positive constants.

III. MAIN RESULTS

A. Distributed Optimization Algorithm Design

A zeroth-order fast distributed optimization algorithm to the problem (6) is proposed as follows:

$$x_i(t) = x_i(0) - \sum_{j=1}^{N} a_{ij} \left(\zeta_i(t) - \zeta_j(t)\right), \tag{7a}$$

$$y_i(t) = x_i(t) + \alpha(t) \left(x_i(t) - x_i(t-1)\right), \tag{7b}$$

$$\begin{aligned} \zeta_i(t+1) &= \zeta_i(t) \\ &\quad + \beta \sum_{j=1}^{N} a_{ij} \left(\hat{\nabla}_{\delta_i(t)} f_i(y_i(t)) - \hat{\nabla}_{\delta_j(t)} f_j(y_j(t))\right), \end{aligned} \tag{7c}$$

where $x_i(t) \in \mathbb{R}$ is the decision variable of generator i, $y_i(t) \in \mathbb{R}$ and $\zeta_i(t) \in \mathbb{R}$ are the auxiliary variables, $i = 1, 2, \cdots, N$, and $\alpha(t)$ and β are the step size parameters. Moreover, $\hat{\nabla}_{\delta_i(t)} f_i(y_i(t))$ is the deterministic gradient estimator defined in (1) with the exploration step size $\delta_i(t)$.

Define $\boldsymbol{x} = [x_1, \cdots, x_N]^T$, $\boldsymbol{\zeta} = [\zeta_1, \cdots, \zeta_N]^T$, $\boldsymbol{y} = [y_1, \cdots, y_N]$, $\boldsymbol{\delta} = [\delta_1, \cdots, \delta_N]$, and $\hat{\nabla} f_{\boldsymbol{\delta}}(\boldsymbol{y}) = [\hat{\nabla} f_{\delta_1}(y_1), \cdots, \hat{\nabla} f_{\delta_N}(y_N)]^T$. Then the algorithm (7) can be written into the compact form

$$\boldsymbol{x}(t) = \boldsymbol{x}(0) - L\boldsymbol{\zeta}(t), \tag{8a}$$

$$\boldsymbol{y}(t) = \boldsymbol{x}(t) + \alpha(t)(\boldsymbol{x}(t) - \boldsymbol{x}(t-1)), \tag{8b}$$

$$\boldsymbol{\zeta}(t+1) = \boldsymbol{\zeta}(t) + \beta L \hat{\nabla} f_{\boldsymbol{\delta}(t)}(\boldsymbol{y}(t)). \tag{8c}$$

Proposition 1: For the algorithm (8), the equality constraint to the problem (6) is always satisfied if $\sum_{i=1}^{N} x_i(0) = D$.

Proof: Noting that $\mathbf{1}^T L = 0$ and premultiplying both sides of (8a) by $\mathbf{1}^T$, we get that $\mathbf{1}^T \boldsymbol{x}(t) = \mathbf{1}^T \boldsymbol{x}(0) = D$, $\forall t > 0$, which implies that Proposition 1 holds. ■

Remark 1: By virtue of the communication topology structure, the algorithm (8) inherently guarantees the satisfaction of the equation constraints. This inherent characteristic sets it apart significantly from the approaches outlined in [23] and [24]. Moreover, motivated by [25], the algorithm (8) employs the momentum method to achieve a more accurate descending direction than the approach presented in [26]. Furthermore, considering the unknown gradient information, the zeroth-order scheme is utilized in the gradient update of the algorithm (8).

B. Convergence Analysis

Now we analyze the convergence of the algorithm.

Theorem 1: Suppose that Assumption 1 holds. Then the algorithm (8) solves the problem (6) at a linear-speedup convergence rate, if $0 < \beta < 1/(l_2 \|L\|^2)$ and $\alpha(t), \delta_i(t) \in (0, \hat{\epsilon}^{t/2}]$, where $\hat{\epsilon} \in (0,1)$. Moreover, there holds

$$\begin{aligned} f(\boldsymbol{x}(t)) - f(\boldsymbol{x}^*) \leq{}& (1-\varepsilon)^t \left(f(\boldsymbol{x}(0)) - f(\boldsymbol{x}^*)\right) \\ &+ (m_1 M + m_2)\phi(\varepsilon, \hat{\varepsilon}, \tilde{\varepsilon}), \end{aligned} \tag{9}$$

where m_1, m_2, $M > 0$ are constants, and

$$\phi(\varepsilon, \hat{\varepsilon}, \tilde{\varepsilon}) = \begin{cases} \dfrac{(1-\varepsilon)^t}{1-\varepsilon-\hat{\varepsilon}}, & \text{if } 1-\varepsilon > \hat{\varepsilon}, \\ \dfrac{\hat{\varepsilon}^t}{\hat{\varepsilon}-1+\varepsilon}, & \text{if } 1-\varepsilon < \hat{\varepsilon}, \\ \dfrac{\tilde{\varepsilon}^t}{\tilde{\varepsilon}-\hat{\varepsilon}}, & \text{if } 1-\varepsilon = \hat{\varepsilon}, \end{cases} \tag{10}$$

with $\tilde{\varepsilon} \in (1-\varepsilon, 1)$.

Proof: From (8a) and (8c), we have

$$\boldsymbol{x}(t+1) - \boldsymbol{x}(t) = -L(\boldsymbol{\zeta}(t+1) - \boldsymbol{\zeta}(t)) = -\beta L^T L \hat{\nabla} f_{\boldsymbol{\delta}(t)}(\boldsymbol{y}(t)).$$

Combined with the fact $\nabla^2 f(\boldsymbol{x}) \leq l_2 I$, we can obtain

$$\begin{aligned} & f(\boldsymbol{x}(t+1)) - f(\boldsymbol{x}(t)) \\ \leq{}& \nabla f(\boldsymbol{x}(t))^T (\boldsymbol{x}(t+1) - \boldsymbol{x}(t)) \\ &+ \frac{l_2}{2} (\boldsymbol{x}(t+1) - \boldsymbol{x}(t))^T (\boldsymbol{x}(t+1) - \boldsymbol{x}(t)) \\ \leq{}& -\beta (L \nabla f(\boldsymbol{x}(t)))^T (L \hat{\nabla} f_{\boldsymbol{\delta}(t)}(\boldsymbol{y}(t))) \\ &+ \frac{l_2 \beta^2 \|L\|^2}{2} \left\| L \hat{\nabla} f_{\boldsymbol{\delta}(t)}(\boldsymbol{y}(t)) \right\|^2 \\ ={}& -\beta (L \nabla f(\boldsymbol{x}(t)) - L \hat{\nabla} f_{\boldsymbol{\delta}(t)}(\boldsymbol{y}(t)))^T \left(L \hat{\nabla} f_{\boldsymbol{\delta}(t)}(\boldsymbol{y}(t))\right) \\ &- (\beta - \frac{l_2 \beta^2 \|L\|^2}{2}) \left\| L \hat{\nabla} f_{\boldsymbol{\delta}(t)}(\boldsymbol{y}(t)) \right\|^2 \\ \leq{}& \frac{\beta}{2} \left\| L \nabla f(\boldsymbol{x}(t)) - L \hat{\nabla} f_{\boldsymbol{\delta}(t)}(\boldsymbol{y}(t)) \right\|^2 \\ &- \left(\frac{\beta}{2} - \frac{l_2 \beta^2 \|L\|^2}{2}\right) \left\| L \hat{\nabla} f_{\boldsymbol{\delta}(t)}(\boldsymbol{y}(t)) \right\|^2, \end{aligned} \tag{11}$$

where the Young's inequality is utilized in the last inequality. For the first term in the last inequality of (11), it follows from Lemma 2 and (8b) that

$$\begin{aligned} & \frac{\beta}{2} \left\| L \nabla f(\boldsymbol{x}(t)) - L \hat{\nabla} f_{\boldsymbol{\delta}(t)}(\boldsymbol{y}(t)) \right\|^2 \\ ={}& \frac{\beta}{2} \left\| L \nabla f(\boldsymbol{x}(t)) - L \nabla f(\boldsymbol{y}(t)) + L \nabla f(\boldsymbol{y}(t)) - L \hat{\nabla} f_{\boldsymbol{\delta}(t)}(\boldsymbol{y}(t)) \right\|^2 \\ \leq{}& \beta l_2^2 \|L\|^2 \alpha^2(t) \|\boldsymbol{x}(t) - \boldsymbol{x}(t-1)\|^2 + \frac{N l_2^2 \beta \|L\|^2}{4} \delta^2(t), \end{aligned} \tag{12}$$

with $\delta(t) = \max_{i=1,\cdots,N}\{\delta_i(t)\}$. For the second term in the last inequality of (11), we have

$$-\left\| L \hat{\nabla} f_{\boldsymbol{\delta}(t)}(\boldsymbol{y}(t)) \right\|^2$$

$$
\begin{aligned}
&= -\left\|L\hat{\nabla}f_{\boldsymbol{\delta}(t)}(\boldsymbol{y}(t)) - L\nabla f(\boldsymbol{y}(t)) + L\nabla f(\boldsymbol{y}(t))\right.\\
&\quad \left. -L\nabla f(\boldsymbol{x}(t)) + L\nabla f(\boldsymbol{x}(t))\right\|^2\\
&\leq \left\|L\hat{\nabla}f_{\boldsymbol{\delta}(t)}(\boldsymbol{y}(t)) - L\nabla f(\boldsymbol{y}(t))\right\|^2 - \frac{1}{4}\|L\nabla f(\boldsymbol{x}(t))\|^2\\
&\quad +\frac{1}{2}\|L\nabla f(\boldsymbol{y}(t)) - L\nabla f(\boldsymbol{x}(t))\|^2\\
&\leq \frac{Nl_2^2\|L\|^2}{4}\delta^2(t) + \frac{l_2^2\|L\|^2}{2}\alpha^2(t)\|\boldsymbol{x}(t) - \boldsymbol{x}(t-1)\|^2\\
&\quad -\frac{1}{4}\|L\nabla f(\boldsymbol{x}(t))\|^2, \qquad (13)
\end{aligned}
$$

where the inequality $-\|a+b+c\|^2 \leq \|a\|^2 + \|b\|^2/2 - \|c\|^2/4$ is used in the first inequality of (13).

Substituting (12) and (13) into (11) gives rise to

$$
\begin{aligned}
&f(\boldsymbol{x}(t{+}1)){-}f(\boldsymbol{x}(t))\\
&\leq m_1\alpha^2(t)\|\boldsymbol{x}(t){-}\boldsymbol{x}(t{-}1)\|^2{+}m_2\delta^2(t)\\
&\quad -\frac{\beta{-}l_2\|L\|^2\beta^2}{8}\|L\nabla f(\boldsymbol{x}(t))\|^2, \qquad (14)
\end{aligned}
$$

where $m_1 = \beta l_2^2\|L\|^2\left(5-\beta l_2\|L\|^2\right)/4$ and $m_2 = \beta N l_2^2\|L\|^2\left(3-\beta l_2\|L\|^2\right)/8$. Then, we get the following inequality using the convex property of $f(\boldsymbol{x})$:

$$
\begin{aligned}
f(\boldsymbol{x}(t)) - f(\boldsymbol{x}^*) &\leq \nabla f(\boldsymbol{x}(t))^T(\boldsymbol{x}(t) - \boldsymbol{x}^*)\\
&\leq -\nabla f(\boldsymbol{x}(t))^T L(\boldsymbol{\zeta}(t) - \boldsymbol{\zeta}^*), \qquad (15)
\end{aligned}
$$

where $\boldsymbol{\zeta}^* \in \mathbb{R}^N$ satisfies $\boldsymbol{x}^* = \boldsymbol{x}(0) - L\boldsymbol{\zeta}^*$. Define a set of the unit orthogonal basis vectors $\boldsymbol{u}_1, \cdots, \boldsymbol{u}_N \in \mathbb{R}^N$, where $\boldsymbol{u}_1$ satisfies $L\boldsymbol{u}_1 = 0$ and $\boldsymbol{u}_1^T\boldsymbol{u}_1 = 1$. Based on this, there exist real numbers $c_1, \cdots, c_N$ at step t such that

$$
\boldsymbol{\zeta}(t) - \boldsymbol{\zeta}^* = \sum_{j=1}^{N} c_i(t)\boldsymbol{u}_i. \qquad (16)
$$

It follows from (15) that

$$
\begin{aligned}
f(\boldsymbol{x}(t)) - f(\boldsymbol{x}^*) &\leq -\nabla f(\boldsymbol{x}(t))^T L\sum_{j=2}^{N} c_i(t)\boldsymbol{u}_i\\
&\leq \|\nabla f(\boldsymbol{x}(t))^T L\|\left\|\sum_{j=2}^{N} c_i(t)\boldsymbol{u}_i\right\|. \qquad (17)
\end{aligned}
$$

Using the Lagrangian multiplier method, we can know that $\nabla f(\boldsymbol{x}^*) = -\lambda^*\mathbf{1}$ with λ^* being the optimal Lagrangian multiplier, which indicates $L\nabla f(\boldsymbol{x}^*) = 0$. Thus, it follows from the condition $\nabla^2 f(\boldsymbol{x}) \geq l_1 I$ that

$$
\begin{aligned}
f(\boldsymbol{x}(t)){-}f(\boldsymbol{x}^*) &\geq \nabla f(\boldsymbol{x}^*)^T(\boldsymbol{x}(t){-}\boldsymbol{x}^*){+}\frac{l_1}{2}\left\|\boldsymbol{x}(t){-}\boldsymbol{x}^*\right\|^2\\
&\geq -\nabla f(\boldsymbol{x}^*)^T L(\boldsymbol{\zeta}(t){-}\boldsymbol{\zeta}^*){+}\frac{l_1}{2}\left\|\boldsymbol{x}(t){-}\boldsymbol{x}^*\right\|^2\\
&\geq \frac{l_1}{2}\left\|\boldsymbol{x}(t){-}\boldsymbol{x}^*\right\|^2. \qquad (18)
\end{aligned}
$$

Moreover, from (16) and Lemma 1, we can get

$$
\begin{aligned}
f(\boldsymbol{x}(t)){-}f(\boldsymbol{x}^*) &\geq \frac{l_1}{2}\left\|L\sum_{j=2}^{N} c_i(t)\boldsymbol{u}_i\right\|^2\\
&\geq \frac{l_1}{2}\lambda_2(L^TL)\left\|\sum_{j=2}^{N} c_i(t)\boldsymbol{u}_i\right\|^2, \qquad (19)
\end{aligned}
$$

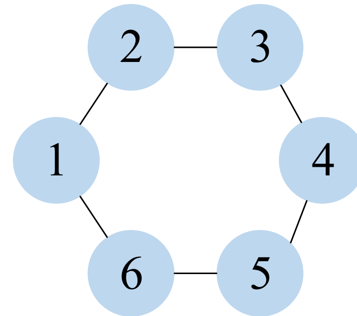

Fig. 1. Communication graph $\mathcal{G}$ with 6 generators.

where $\lambda_2(L^TL)$ represents the second smallest eigenvalue of L^TL. It can be deduced from (17) and (19) that

$$
\begin{aligned}
&\left\|\nabla f(\boldsymbol{x}(t))^T L\right\|^2\left\|\sum_{j=2}^{N} c_i(t)\boldsymbol{u}_i\right\|^2\\
&\geq \left(f(\boldsymbol{x}(t)){-}f(\boldsymbol{x}^*)\right)^2\\
&\geq \frac{l_1}{2}\lambda_2(L^TL)\left\|\sum_{j=2}^{N} c_i(t)\boldsymbol{u}_i\right\|^2\left(f(\boldsymbol{x}(t)){-}f(\boldsymbol{x}^*)\right),
\end{aligned}
$$

which indicates

$$
-\|L\nabla f(\boldsymbol{x}(t))\|^2 \leq -\frac{l_1}{2}\lambda_2(L^TL)(f(\boldsymbol{x}(t)){-}f(\boldsymbol{x}^*)). \quad (20)
$$

Substituting (20) into (14) produces

$$
\begin{aligned}
f(\boldsymbol{x}(t+1)){-}f(\boldsymbol{x}^*) &\leq (1{-}\epsilon)(f(\boldsymbol{x}(t)){-}f(\boldsymbol{x}^*)){+}m_2\delta^2(t)\\
&\quad {+}m_1\alpha^2(t)\|\boldsymbol{x}(t){-}\boldsymbol{x}(t{-}1)\|^2, \quad (21)
\end{aligned}
$$

where $\epsilon = (\beta{-}l_2\|L\|^2\beta^2)l_1\lambda_2(L^TL)/16$. From Proposition 1, we can know that $x_i(t)$ is bounded, thus let $\|\boldsymbol{x}(t) - \boldsymbol{x}(t-1)\|^2 \leq M$, $\forall t > 0$. We can further write (21) as

$$
\begin{aligned}
f(\boldsymbol{x}(t+1)){-}f(\boldsymbol{x}^*) &\leq (1{-}\epsilon)(f(\boldsymbol{x}(t)){-}f(\boldsymbol{x}^*)){+}m_2\delta^2(t)\\
&\quad {+}m_1 M\alpha^2(t). \qquad (22)
\end{aligned}
$$

It is easy to get

$$
\begin{aligned}
&f(\boldsymbol{x}(t+1)){-}f(\boldsymbol{x}^*) \leq (1{-}\epsilon)^{t+1}\left(f(\boldsymbol{x}(0)){-}f(\boldsymbol{x}^*)\right)\\
&\quad + \sum_{\tau=0}^{t}(1-\varepsilon)^{\tau}\left(m_1 M\alpha^2(t-\tau) + m_2\delta^2(t-\tau)\right). \quad (23)
\end{aligned}
$$

With the parameters designed in the theorem, we can obtain (9) from (23), which completes the proof. ■

Remark 2: According to Theorem 1, we can know that the algorithm (8) can fast converge to the exact optimal solution of the problem (6) with the appropriate parameter design.

IV. SIMULATION EXAMPLES

Numerical examples are presented in this section to verify the theoretical results.

We validate the effectiveness of the designed algorithm (8) on the IEEE 30-bus system with 6 generators and 30 buses. The communication interaction topology $\mathcal{G}$ satisfying

TABLE I

Coefficients Setting for the Economic Cost

G	$a_i[\$/MW^2h]$	$b_i[\$/MWh]$	$c_i[\$/MWh]$	$d_i[\$/MW^2h]$
1	0.00375	2.0	0	0.00005437
2	0.0175	1.75	0	0.00003421
3	0.0625	1.0	0	0.00004213
4	0.00834	3.25	0	0.00002812
5	0.025	3.0	0	0.00003523
6	0.025	3.0	0	0.00003352

TABLE II

Coefficients Setting for the Environmental Pollution

G	$\tilde{a}_i[\text{ton}/MW^2]$	$\tilde{b}_i[\text{ton}/MW]$	$\tilde{c}_i[\text{ton}/MW]$
1	0.0101	-0.0277	0
2	0.0156	-0.0302	0
3	0.0124	-0.0254	0
4	0.0245	-0.0177	0
5	0.0125	-0.0271	0
6	0.0532	-0.0045	0

Assumption 1 is shown in Fig. 1. The involved coefficients of the economic cost and the environmental pollution for generators are listed in Table I and Table II, respectively. Moreover, set the weight coefficients $\eta_1 = 0.7$, $\eta_2 = 0.3$ and the total demand $D = 300MW$. The initial states of generators are set as $x_1(0) = 80MW$, $x_2(0) = 120MW$, $x_3(0) = 5MW$, $x_4(0) = 45MW$, $x_5(0) = 35MW$, and $x_6(0) = 15MW$, which satisfy $\sum_{i=1}^{6} x_i(0) = D$. The exploration parameters are uniformly set as $\delta_i(t) = 0.8^t$ for convenience, and let the step sizes $\alpha(t) = 0.568^{0.6t}$ and $\beta = 0.05$. To demonstrate the validity and flexibility of the proposed algorithm, some simulations over the undirected graph depicted in Fig. 1 will be performed in Example 1 and Example 2, respectively.

Example 1 (Convergence verification): For the distribution power grid, there is the line loss during the actual transmission process. With the aforementioned parameter settings, the simulation results of the algorithm (8) are presented in Figs. 2–4. Specifically, it can be seen from Fig. 2 that the supply power of each generator can converge to the optimal value, respectively, i.e., $x_1^* = 149.5952MW$, $x_2^* = 55.4165MW$, $x_3^* = 25.2910MW$, $x_4^* = 31.2435MW$, $x_5^* = 23.5757MW$, and $x_6^* = 14.8782MW$. The combined cost also reaches the optimum value $f^* = 686.5190\$$ as shown in Fig. 3. It is noted that the total power supply of the generators is always equal to the total demand D during the algorithm operation as indicated in Fig. 4.

Example 2 (Plug-and-play capability): To further present the flexibility of the designed algorithm, we consider the plug and play of generator 6 in this example. Assume that generator 6 loses connection at iteration step $t = 150$. Let anyone of its neighbors takes on its current supply. Then generator 6 returns to the power system with $x_6(0) = 0MW$ at $t = 300$. The simulation results are displayed in Figs. 5 and 6, respectively. As can be seen from Fig. 5, the remaining five generators begin to seek the new optimal solution at $t = 150$. After generator 6 returns to the system at $t = 300$, all generators can anew reach the previous optimal solution quickly. Accordingly, the cost reaches the optimal value after the system experiences two fluctuations at $t = 150$ and $t = 300$ as shown in Fig. 6.

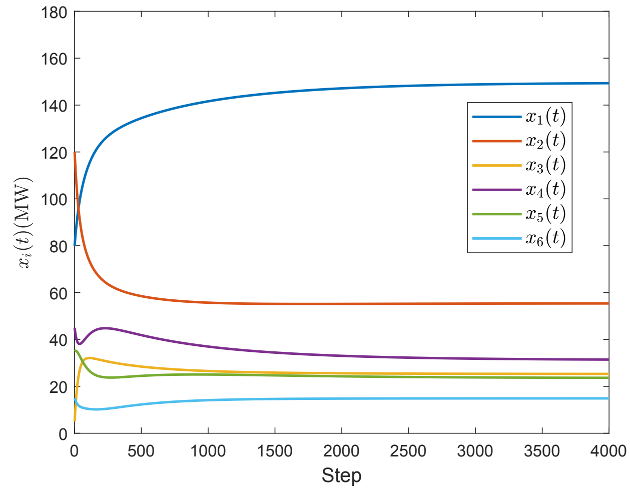

Fig. 2. The evolution of the power supplies $x_1, x_2, \cdots, x_6$ under the algorithm (8) in Example 1.

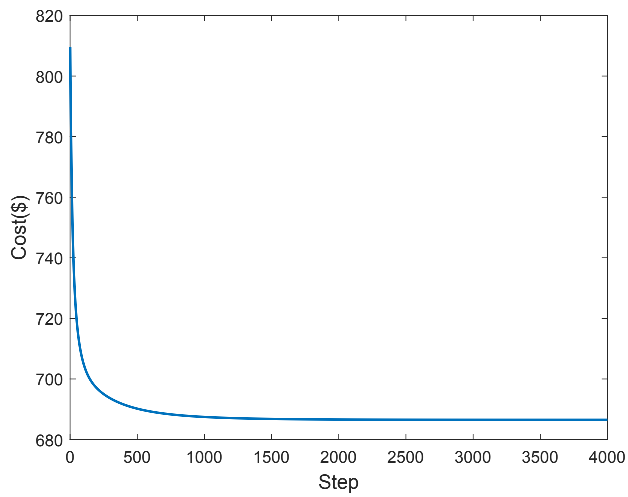

Fig. 3. The combined cost evolution f under the algorithm (8) in Example 1.

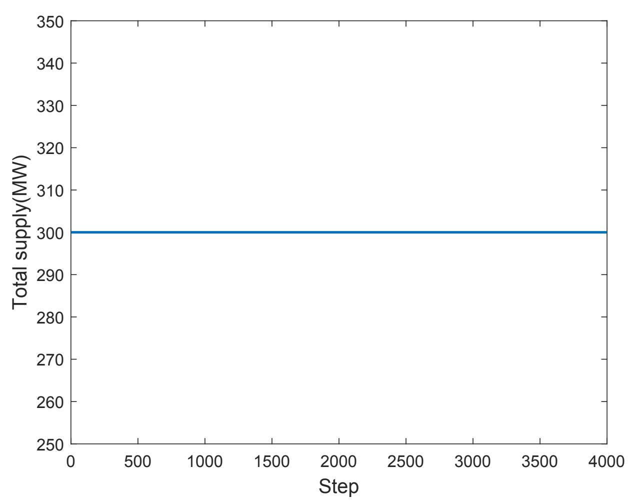

Fig. 4. The evolution of the total power supply $\sum_{i=1}^{6} x_i$ under the algorithm (8) in Example 1.

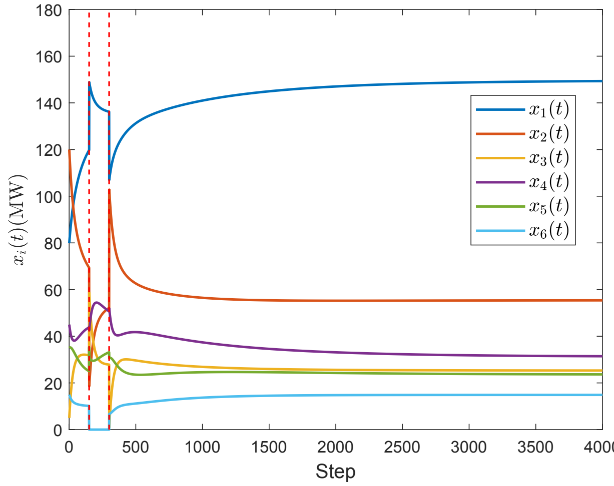

Fig. 5. The evolution of the power supplies $x_1, x_2, \cdots, x_6$ under the algorithm (8) in Example 2.

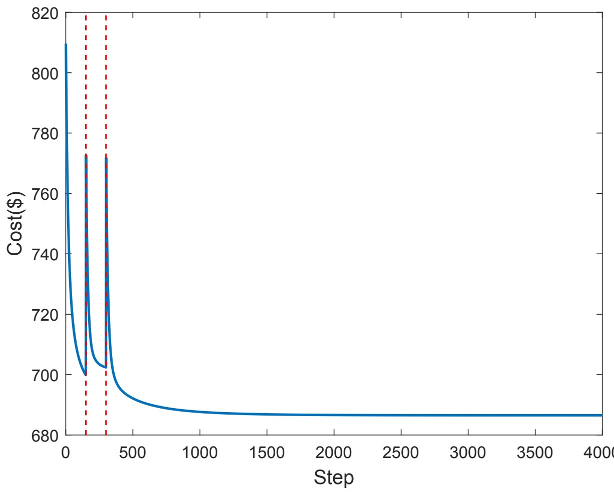

Fig. 6. The cost evolution under the algorithm (8) in Example 2.

V. CONCLUSION

In this paper, a multi-objective DRA problem has been constructed, which takes into account the economic cost and the environmental pollution simultaneously. For the established optimization problem, a zeroth-order accelerated distributed optimization algorithm has been proposed. For the algorithm design, the momentum method has been employed to accelerate the convergence speed through estimating a more accurate descent direction. Moreover, the zeroth-order scheme has been utilized to address the unknown gradient information. It has been shown that the proposed algorithm can reach the linear-speedup convergence rate and can always satisfy the equality constraint. Future work will be to design an accelerated DRA algorithm over directed graphs subject to multiple constraints.

REFERENCES

[1] T. Yang, X. Yi, J. Wu, Y. Yuan, D. Wu, Z. Meng, Y. Hong, H. Wang, Z. Lin, and K. H. Johansson, "A survey of distributed optimization," *Annu. Rev. Control*, vol. 47, pp. 278–305, 2019.

[2] G. Wen, G. Hu, J. Hu, X. Shi, and G. Chen, "Frequency regulation of source-grid-load systems: A compound control strategy," *IEEE Trans. Ind. Electron.*, vol. 12, no. 1, pp. 69–78, Feb. 2016.

[3] N. Jain, S. N. Singh, and S. C. Srivastava, "A generalized approach for DG planning and viability analysis under market scenario," *IEEE Trans. Ind. Electron.*, vol. 60, no. 11, pp. 5075–5085, Nov. 2013.

[4] A. Chaouachi, R. M. Kamel, R. Andoulsi, and K. Nagasaka, "Multi-objective intelligent energy management for a microgrid," *IEEE Trans. Ind. Electron.*, vol. 60, no. 4, pp. 1688–1699, Apr. 2013.

[5] G. Qian, Z. Li, C. He, X. Li, and X. Ding, "Power allocation schemes based on deep learning for distributed antenna systems," *IEEE Access*, vol. 8, pp. 31245–31253, Feb. 2020.

[6] G. Wen, X. Yu, Z.-W. Liu, and W. Yu, "Adaptive consensus-based robust strategy for economic dispatch of smart grids subject to communication uncertainties," *IEEE Trans. Ind. Electron.*, vol. 14, no. 6, pp. 2484–2496, Jun. 2018.

[7] X. Zeng, J. Lei, and J. Chen, "Dynamical primal-dual accelerated method with applications to network optimization," *IEEE Trans. Autom. Control*, vol. 68, no. 3, pp. 1760–1767, Mar. 2023.

[8] R. Deng, G. Xiao, R. Lu, and J. Chen, "Fast distributed demand response with spatially and temporally coupled constraints in smart grid," *IEEE Trans. Ind. Electron.*, vol. 11, no. 6, pp. 1597–1606, Dec. 2015.

[9] W. Shi, Q. Ling, G. Wu, and W. Yin, "EXTRA: An exact first-order algorithm for decentralized consensus optimization," *SIAM J. Optim.*, vol. 25, no. 2, pp. 944–966, May 2015.

[10] R. Xin and U. A. Khan, "Distributed heavy-ball: A generalization and acceleration of first-order methods with gradient tracking," *IEEE Trans. Autom. Control*, vol. 65, no. 6, pp. 2627–2633, Jun. 2020.

[11] G. Qu and N. Li, "Accelerated distributed Nesterov gradient descent," *IEEE Trans. Autom. Control*, vol. 65, no. 6, pp. 2566–2581, Jun. 2020.

[12] A. Nedić, A. Olshevsky, and W. Shi, "Achieving geometric convergence for distributed optimization over time-varying graphs," *SIAM J. Optim.*, vol. 27, no. 4, pp. 2597–2633, Dec. 2017.

[13] Y. Tang, J. Zhang, and N. Li, "Distributed zero-order algorithms for nonconvex multiagent optimization," *IEEE Trans. Control Netw. Syst.*, vol. 8, no. 1, pp. 269–281, Mar. 2021.

[14] J. C. Duchi, M. I. Jordan, M. J. Wainwright, and A. Wibisono, "Optimal rates for zero-order convex optimization: The power of two function evaluations," *IEEE Trans. Inf. Theory*, vol. 61, no. 5, pp. 2788–2806, May 2015.

[15] Y. Nesterov and V. Spokoiny, "Random gradient-free minimization of convex functions," *Found. Comput. Math.*, vol. 17, no. 2, pp. 527–566, Apr. 2017.

[16] S. Liu, J. Chen, P.-Y. Chen, and A. O. Hero, "Zeroth-order online alternating direction method of multipliers: Convergence analysis and applications," *arXiv:1710.07804v2*, 2018.

[17] A. Agarwal, O. Dekel, and L. Xiao, "Optimal algorithms for online convex optimization with multi-point bandit feedback," in *Proc. 23rd Conf. Learning Theory*, Haifa, Israel, Jun. 2010, pp. 28–40.

[18] S. Yang, S. Tan, and J. Xu, "Consensus based approach for economic dispatch problem in a smart grid," *IEEE Trans. Power Syst.*, vol. 28, no. 4, pp. 4416–4426, Apr. 2013.

[19] A. J. Wood, B. F. Wollenberg, and G. B. Sheblé, *Power Generation, Operation, and Control*, 2nd ed. New York, NY, USA: Wiley, 1996.

[20] M. Hamdi, M. Chaoui, L. Idoumghar, and A. Kachouri, "Coordinated consensus for smart grid economic environmental power dispatch with dynamic communication network," *IET Gener. Transm. Distrib.*, vol. 12, no. 11, pp. 2603–2613, Apr. 2018.

[21] B. Shi, L. Yan, and W. Wu, "Multi-objective optimization for combined heat and power economic dispatch with power transmission loss and emission reduction," *Energy*, vol. 56, no. 1, pp. 135–143, Jul. 2013.

[22] K. Miettinen, *Nonlinear Multiobjective Optimization*. New York, NY, USA: Springer, 1998.

[23] W. Li, Y. Liu, H. Liang, and Y. Shen, "A new distributed energy management strategy for smart grid with stochastic wind power," *IEEE Trans. Ind. Electron.*, vol. 68, no. 2, pp. 1311–1321, Feb. 2021.

[24] L.-N. Liu and G.-H. Yang, "Distributed optimal economic environmental dispatch for microgrids over time-varying directed communication graph," *IEEE Trans. Netw. Sci. Eng.*, vol. 8, no. 2, pp. 1913–1924, Apr.-Jun. 2021.

[25] J. Gao, X. Liu, Y.-H. Dai, Y. Huang, and P. Yang, "A family of distributed momentum methods over directed graphs with linear convergence," *IEEE Trans. Autom. Control*, vol. 68, no. 2, pp. 1085–1092, Feb. 2023.

[26] J. Zhou, Y. Lv, C. Wen, and G. Wen, "Solving specified-time distributed optimization problem via sampled-data-based algorithm," *IEEE Trans. Netw. Sci. Eng.*, vol. 9, no. 4, pp. 135–143, Jul.-Aug. 2022.