Disturbance-resilient Distributed Resource Allocation over Stochastic Networks using Uncoordinated Stepsizes

Tie Ding, Shanying Zhu, IEEE Member, Cailian Chen, IEEE Member, Xinping Guan, IEEE Fellow

Abstract—This paper studies distributed resource allocation problem in multi-agent systems, where all the agents cooperatively minimize the sum of their cost functions with global resource constraints over stochastic communication networks. This problem arises from many practical domains such as economic dispatch in smart grid, task assignment and power allocation in robotic control. Most of existing works cannot converge to the optimal solution if states deviate from feasible region due to disturbance caused by environmental noise, misoperation, malicious attack, etc. To solve this problem, we propose a distributed deviation-tracking resource allocation algorithm and prove that it linearly converges to the optimal solution with constant stepsizes. We further explore its resilience properties of the proposed algorithm. Most importantly, the algorithm still converges to the optimal solution under the disturbance injection and random communication failure. In order to improve the convergence rate, the optimal stepsizes for the fastest convergence rate are established. We also prove the algorithm converges linearly to the optimal solution in mean square even with uncoordinated stepsizes, i.e., agents are allowed to employ different stepsizes. Simulations are provided to verify the theoretical results.

Index Terms—Distributed resource allocation, Stochastic network, Deviation tracking, Resilience to disturbance

I. INTRODUCTION

Recently, distributed resource allocation problem has attracted much attention due to its wide application in various practical problems, such as economic dispatch in energy network [1]–[3], channel allocation in wireless communication [4], [5] and computing resource assignment in edge computing [6]. In resource allocation problems, centralized methods require an entity to collect information of agents and distribute the strategies of resource allocation or task assignment to all the agents [7]. It suffers from issues of high requirement of synchronization, heavy cost of long-distance communication and poor scalability. Distributed resource allocation approaches avoid long-distance communication and make the network more scalable [8]. Since each node only has local knowledge, it requires a reliable communication network to achieve global optimization. Due to the vulnerability nature of wireless communications, the network is easily affected by potential attack and environmental noise. This may lead to a stochastic communication network suffering from random failure, which results in information island, inaccurate estimation of the optimal solution and, eventually, inexact and stochastic convergence result. Therefore, it is significant to design proper distributed algorithms to obtain the optimal strategies effectively and simultaneously alleviate the stochasticity of convergence result resulting from link failures.

Distributed resource allocation problems are actually the dual problems of distributed consensus optimization, which requires all agents’ states to be equal while the optimal point of distributed resource allocation problems is obtained when achieving consensus on marginal costs. Due to this property, [9] proposes decentralized resource allocation algorithm that adopts weighted gradient method to ensure the consensus on gradients. Extension of the result [9] to random gradients, the authors [10] design a random coordinate descent algorithm based on weighted gradient method and prove that it converges to the optimal solution linearly in probability. However, both [9] and [10] are only suitable for fixed communication network and does not consider stochastic network with communication failure. The authors of [11] extend the algorithm of [9] to time-varying networks and prove that it converges to the optimal solution in quadratic time. But [11], as well as [9], [10], requires the states to be kept feasible through all iterations. If the state is disturbed by noise or malicious attacks at any moment, which causes infeasible states, a derivation from the optimal point occurs inevitably.

To deal with infeasible states, [12] proposes an initialization-free distributed algorithm to solve optimal resource allocation problems with local constraints. But it is a continuous-time algorithm requiring infinite bandwidth to support the data exchange between agents and the convergence rate is far from optimal. Additionally, continuous-time algorithms cannot deal with stochastic communication network because it is difficult to define derivative with stochastic variables. Ref. [13] proposes a distributed algorithm over dynamic networks, considering the uncertainty of local parameters. It converges to the optimal point but a decaying stepsize is needed, which results in a slower convergence rate. A decaying stepsize is also adopted in [14]. Using constant stepsize, [15] proposes a distributed resource allocation algorithm with dual splitting approach (DuSPA), which ensures that both primary and dual states converge to the optimal point. But it can only be used in fixed network. Extending primary-dual methods to time-varying networks, the authors of [16], [17] propose distributed resource allocation algorithms which converge linearly to the optimal solution with constant stepsizes. Both [16] and [17] assume that the network is jointly strongly connected, i.e., each communication link should have bounded intercommunication interval. For stochastic networks, however, we cannot determine the bound of communication interval, especially when the possibility of packet loss is high. Moreover, these aforementioned works all adopt identical stepsizes for all agents. It is difficult for all agents to realize the consensus of stepsizes in distributed

The authors are with the Department of Automation, Shanghai Jiao Tong University, Shanghai 200240, China, and also with Key Laboratory of System Control and Information Processing, Ministry of Education of China, Shanghai 200240, China.

This work is supported in part by National Key Research and Development Program of China (No. 2016YFB0901900), NSF of China under the grants 61922058, 61633017, and 61521063, and NSF of Shanghai Municipality of China under grant 18ZR1419900.
manner due to random communication failure.

As one typical application of distributed resource allocation, economic dispatch problems in smart grids, where a group of microgrids cooperatively minimize the cost of generation subject to the balance between supply and demand, also gain lots of attention. The authors of [13] propose an incremental cost consensus algorithm that solves economic dispatch problem in a distributed way. However, it requires a center entity to maintain the balance between supply and demand. To eliminate the centralized node, a consensus + innovation problem in a distributed way. However, it requires a center entity to maintain the balance between supply and demand. To eliminate the centralized node, a consensus + innovation algorithm works for general functions with Lipschitz gradients and strong convexity. Moreover, compared with [16], [17], we relax the assumption of joint connectivity of networks over certain bounded intercommunication interval that used in to connectivity in mean.

1) We propose a disturbance-resilient distributed algorithm to solve distributed resource allocation problems over stochastic networks with random communication failure and stochastic disturbance on states. This algorithm is a combination of the deviation-tracking technique and weighted gradient scheme. Different from most economic dispatch algorithms that only are suitable for strictly quadratic cost functions of generation, which limits scope of application. Moreover, these aforementioned works do not consider the communication failure, which may be caused by limited transmission energy, environmental noise, malicious attack, etc. Communication failure may cause the power system to be operated in a non-optimal condition, which will greatly increase the productivity cost [21].

We aim to solve distributed resource allocation problem over stochastic networks where communication links randomly fail and the states are injected by random disturbance. In this paper, a disturbance-resilient distributed algorithm targeting the communication failures is proposed with guaranteed linear convergence to the optimal solution.

The contributions of this paper are shown as follows.

1) We propose a disturbance-resilient distributed algorithm to solve distributed resource allocation problems over stochastic networks with random communication failure and stochastic disturbance on states. This algorithm is a combination of the deviation-tracking technique and weighted gradient scheme. Different from most economic dispatch algorithms that only are suitable for strictly quadratic cost functions [13], [20], the proposed algorithm works for general functions with Lipschitz gradients and strong convexity. Moreover, compared with [16], [17], we relax the assumption of joint connectivity of networks over certain bounded intercommunication interval that used in to connectivity in mean.

2) We prove that the proposed algorithm converges linearly to the optimal solution in mean square even with random communication failure. A specific upper bound of the stepsizes that ensure convergence is provided. We further explore the resilience properties of the proposed algorithm. Compared with the algorithms proposed in [9], [11], [15], the proposed algorithm converges in mean square to the optimal solution even with disturbance on states while these works cannot reside from disturbance. We provide comparative simulations in Section VII B.

3) Based on the convergence results, we obtain the estimate of the convergence rate. To improve the convergence rate, the optimal stepsizes within the upper bound are established. Furthermore, we prove that the algorithm converges to the optimal solution in mean square even with uncoordinated stepsizes.

Notation: We denote the set of $n$-dimensional vectors and $n \times u$-dimensional matrices by $\mathbb{R}^n$ and $\mathbb{R}^{n \times u}$, respectively.

0, 1 $\in \mathbb{R}^n$ represents the vectors of zeros and ones, respectively. $I \in \mathbb{R}^{n \times n}$ is the identity matrix. $\| \cdot \|_F$ denotes the Frobenius norm for matrices and $\| \cdot \|_2$ represents the Euclidean norm for vectors. $\{ \cdot \}$ means the expectation and for any vector $v \in \mathbb{R}^n$, we define expectation norm $\| v \|_E = \sqrt{\mathbb{E}[\| v \|_2^2]}$.

For any vector sequence $\{ v(k) \}_{k=1}^\infty$, $v(k)$ converges to $v^*$ in mean square if $\lim_{k \to \infty} \| v(k) - v^* \|_E = 0$. $\rho(\cdot)$ denotes the spectral radius of matrices. For any real symmetric matrix $A \in \mathbb{R}^{n \times n}$, let $\{ \lambda_i(A) \}_{i=1}^n$ be the eigenvalues such that $\lambda_n(A) \leq \lambda_{n-1}(A) \leq \ldots \leq \lambda_1(A)$. We denote by $x_i$ the local copy of the global variable $x \in \mathbb{R}^n$ at agent $i$. Its value at time $k$ is denoted by $x_i(k)$. We introduce the stacked matrix $x = [x_1, \ldots, x_n]^T \in \mathbb{R}^{n \times u}$ and define $\bar{x} = \frac{1}{n}x$, $\hat{x} = Lx$, where $L = I - \frac{1}{n}1_n1_n^T$.

II. Problem Formulation

Consider a distributed resource allocation problem in multi-agent systems. For any $i \in \{1, \ldots, n\}$, agent $i$ has its individual cost function $f_i: \mathbb{R}^u \to \mathbb{R}$, which is only known to agent $i$ itself. All the agents cooperate to minimize the sum of their cost functions with specified amount of resource:

$$\min_{x \in \mathbb{R}^{n \times u}} f(x) = \sum_{i=1}^n f_i(x_i) \text{ s.t. } \sum_{i=1}^n x_i = \sum_{i=1}^n d_i,$$

(1)

where $d_i, \forall i \in \{1, \ldots, n\}$ denotes the local demand of resource, which is only known by agent $i$ and does not share with other agents for privacy concern. Here we make the following standard assumptions about the functions $f_i, i \in \{1, \ldots, n\}$.

Assumption 1. For every $i$, $f_i: \mathbb{R}^u \to \mathbb{R}$ is differentiable and has Lipschitz gradient with $\eta_i > 0$.

$$\| \nabla f_i(x_a) - \nabla f_i(x_b) \|_2 \leq \phi_i \| x_a - x_b \|_2, \forall x_a, x_b \in \mathbb{R}^u.$$  (2)

Assumption 2. For every $i$, $f_i: \mathbb{R}^u \to \mathbb{R}$ is strongly convex, i.e., there exists $\phi_i > 0$ such that $\forall x_a, x_b \in \mathbb{R}^n$,

$$(x_a - x_b)^T (\nabla f_i(x_a) - \nabla f_i(x_b)) \geq \eta_i \| x_a - x_b \|_2^2.$$  (3)

We model the topology of the network over which agents communicate with each other as a random graph $G(k) = (V, E(k))$, where $V$ is the set of agents $V = \{1, 2, \ldots, n\}$, $E(k) \subset V \times V$ denotes the edges. Let $W(k) = [w_{ij}(k)]_{n \times n} \in \mathbb{R}^{n \times n}$, where $w_{ij}(k)$ is the weight of the edge $(i, j)$ at time $k$. $w_{ij}(k) > 0$ means that agent $i$ can communicate with agent $j$ at time $k$. Since each agent has local knowledge, we have $w_{ij}(k) > 0, \forall i, k$. All the agents in $N_i(k) = \{j | w_{ij}(k) > 0\}$ are called neighbors of agent $i$. Here, we consider that each communication link is subject to a random failure, that is, for any agent $i$ and $j \in N_i(k)$, we have $P\{w_{ij}(k) > 0\} = \theta_{ij}(k)$ and $P\{w_{ij}(k) = 0\} = 1 - \theta_{ij}(k), 0 < \theta_{ij}(k) < 1$. Specifically, $P\{w_{ij}(k) > 0\} = 1, \forall i$. We assume that if link $(i, j)$ fails, link $(j, i)$ also fails.

Here, we provide a method to determine the weight $w_{ij}(k), \forall i, j, k$. For each time $k$, agent $i$ sends information to agent $j$, along with the weight $w_{ij}(k), j \in N_i$. If the communication link between $i$ and $j$ works, agent $i$ and $j$ will receive $w_{ij}(k)$ and $w_{ij}(k)$, respectively. The weight of the edge $(i, j)$ is set as the smaller one between $w_{ii}(k)$ and $w_{jj}(k)$, i.e.,

$$w_{ij}(k) = \begin{cases} \min\{w_{ii}(k), w_{jj}(k)\} & \text{if link } (i, j) \text{ works} \\ 0 & \text{if link } (i, j) \text{ fails} \end{cases}.$$  (4)


Then, for each agent $i$,

$$w_{ii}(k) = 1 - \sum_{j \in \mathcal{N}_i} w_{ij}(k),$$

and thus $\sum_{j=1}^{n} w_{ij}(k) = 1$ and $\sum_{j=1}^{n} w_{ji}(k) = 1, \forall i, k$.

Based on the above method to obtain the weights, we make the following assumption on the communication network \cite{22, 23}.

**Assumption 3.** The weight matrices \{\(W(k)\)\} are a sequence of i.i.d matrices from some probability space \(F = (\Omega, B, P)\) such that each \(W(k)\) is symmetric, doubly stochastic, i.e., \(\forall k\)

\[
W(k) = W(k)^T, \quad 1^T W(k) = 1^T, \quad W(k) 1 = 1,
\]

and

$$\rho \left( \mathbb{E}\{W(k)\} - \frac{11^T}{n} \right) < 1. \quad \tag{7}$$

Eq.\(7\) is equivalent to that the graph is connected in mean. It should be noted that a connected graph is important for all agents to achieve global optimal solution with only local communication. In \cite{16} and \cite{17}, the graph is assumed to be jointly connected, i.e., for any \(k\), within constant \(B\) intervals, the joint graph \(\bigcup_{k=k-B}^{k} \mathcal{E}(t)\) is connected. In stochastic networks, due to the randomness of communication link failure, we cannot determine a constant \(B\) such that the joint graph is connected with every \(B\) intervals. In this paper, we only require the the graph to be connected in mean, which means that the communication network may be disconnected at each interval. This is different from the assumption of bounded intercommunication interval in \cite{16, 17}.

Let $\lambda_{2e} = \max_{k} \lambda_2(\mathbb{E}\{W(k)\})$, $\lambda_{nn} = \min_{k} \lambda_n(\mathbb{E}\{W(k)\})$. Due to \(7\), we have that $-1 < \lambda_{nn} \leq \lambda_{2e} < 1$. Moreover, let $\lambda_{\nu} = \inf_k \{\lambda_n(W(k))\}$. Since $W(k)$ is doubly stochastic, $w_{ii}(k) > 0$ and $w_{ij}(k) \geq 0$, we have $\lambda_{\nu} > -1$.

### III. ALGORITHM DEVELOPMENT

#### A. Optimal conditions

Before introducing our algorithm, we give the conditions of the optimal solution.

**Proposition 1.** $x^* = [x_1^*, ..., x_n^*]^T \in \mathbb{R}^{n \times u}$ is the optimal solution of problem (1) if $x^*$ satisfies the following two conditions.

(i) \(1^T x = 1 \cdot d,\)

(ii) there exists $\mu^* \in \mathbb{R}^u$ such that $\nabla f(x^*) = 1(\mu^*)^T$.

**Proof.** Define the Lagrange function

$$L(x, \mu) = \sum_{i=1}^{n} f_i(x_i) + \mu \left( \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} d_i \right), \tag{8}$$

where $\mu \in \mathbb{R}^u$ is the Lagrange multiplier. By the KKT conditions \cite{24}, we obtain that

$$\nabla f_i(x_i^*) - \mu^* = 0, \forall i, \quad \sum_{i=1}^{n} x_i^* - \sum_{i=1}^{n} d_i = 0. \tag{9}$$

Then, we obtain that

$$\nabla f(x^*) = [\nabla f_1(x_1^*), ..., \nabla f_n(x_n^*)]^T = 1(\mu^*)^T. \tag{10}$$

Eqs.\(9\) and \(10\) verify the conditions (i) and (ii), respectively, completing the proof.

Proposition \(1\) implies that the optimal point is achieved when all the states are feasible and simultaneously the gradients of all agents’ functions are equal, i.e., achieving consensus on marginal costs while keeping the balance between supply and demand.

#### B. Distributed deviation-tracking method

To solve the problem \(\Pi\), we develop a distributed resource allocation algorithm by combining weighted gradient method and deviation-tracking scheme.

First, we adopt weighted gradient method to ensure that the conditions of Proposition 1, we consider the following iteration.

$$x_i(k + 1) = x_i(k) - \beta \sum_{j=1}^{n} w_{ij}(k)(\nabla f_i(x_i(k)) - \nabla f_j(x_j(k))). \quad \tag{11}$$

This rule is also adopted in other distributed resource allocation problems over fixed networks \cite{9, 10}. For the Condition (i) of Proposition 1, we consider the following iteration.

$$x_i(k + 1) = x_i(k) - \frac{\alpha}{n} \sum_{j=1}^{n} (x_j(k) - d_j). \quad \tag{12}$$

It is clear that if $x_i(k)$ converges, the deviation $\sum_{j=1}^{n} (x_j(k) - d_j)$ converges to 0, which is equivalent to Condition (i) of Proposition 1.

However, it is noted that the variable $\frac{1}{n} \sum_{j=1}^{n} x_j(k) - d_j$ requires all agents’ states at each iteration, which is a global variable that cannot be obtained by each agent. We propose a deviation-tracking method to decentralize the global deviation $\frac{1}{n} \sum_{j=1}^{n} x_j(k) - d_j$ by introducing an auxiliary variable $y_i(k)$ to track it. We attempt to use consensus protocol to ensure $y_i(k)$ tracks $\sum_{j=1}^{n} x_j(k) - d_j$. But based on consensus protocol, the sum of $\sum_{j=1}^{n} y_j(k)$ is constant for all $k$. Now that $\frac{1}{n} \sum_{j=1}^{n} x_j(k)$ is time-varying, so we add a compensation term to track the value $\frac{1}{n} \sum_{j=1}^{n} x_j(k)$. Starting from $y_i(0) = x_i(0) - d_i$, the update rule of agent $i$ is expressed as follows.

$$x_i(k + 1) = x_i(k) - \alpha y_i(k),$$

$$- \beta \sum_{j=1}^{n} w_{ij}(k)(\nabla f_i(x_i(k)) - \nabla f_j(x_j(k)))$$

$$y_i(k + 1) = \sum_{j=1}^{n} w_{ij}(k) y_j(k) + x_i(k + 1) - x_i(k), \quad \tag{13}$$

where $\alpha$ and $\beta$ are constant stepizes. $y_i(k + 1)$ is the auxiliary variable tracking the global deviation $\sum_{j=1}^{n} (x_j(k) - d_j)$.

The algorithm is shown in Algorithm 1 for details. Let

$$\nabla f(x(k)) = [\nabla f_1(x_1(k)), ..., \nabla f_n(x_n(k))]^T \in \mathbb{R}^{n \times u}, \forall k.$$ 

Then, the algorithm can be rewritten in a matrix form:

$$y(0) = x(0) - d, $$

$$x(k + 1) = x(k) - \alpha y(k) - \beta(I - W(k)) \nabla f(x(k)), $$

$$y(k + 1) = W(k)y(k) + x(k + 1) - x(k). $$

To intuitively describe the algorithm of deviation-tracking method, we provide the block diagram of our algorithm. In
Algorithm 1: Distributed Deviation-tracking Resource Allocation Algorithm

Input: $W$, $f_i$, $\alpha$, $\beta$, $d_i$
Output: $x_i$, $y_i$
1: Initialization: Pick any $x_i(0) \in \mathbb{R}^n$, $y_i(0) = x_i(0) - d_i$
2: for $k = 0$ : $\infty$ do
3:   Broadcast $\nabla f_j(x_j(k))$ and $y_i(k)$ to neighbors;
4:   Receive $\nabla f_j(x_j(k))$ and $y_j(k)$, $j \in \mathcal{N}_i$ from neighbors;
5:   Update $x_i(k + 1)$ through
6:   $x_i(k + 1) = x_i(k) - \beta \nabla f_j(x_j(k)) - \sum_{j=1}^{n} w_{ij}(k) \nabla f_j(x_j(k));$
7:   Update $y_i(k + 1)$ through
8:   $y_i(k + 1) = \sum_{j=1}^{n} w_{ij}(k) y_j(k) + x_i(k + 1) - x_i(k);
9: end for

Fig. 1. Block diagram of Algorithm 1.

Algorithm 1, we adopt two feedback loops to ensure the optimal convergence. As shown in Fig. 1, we first use the $(I - W(k))\nabla f_i(x_i(k))$ to guarantee that the states converge with equal gradients because $\nabla f_j(x_j(k)) = \nabla f_j(x_j(k)), \forall i, j \iff (I - W(k))\nabla f_i(x_i(k)) = 0$. Then, we introduce $y_i(k)$ to track the deviation $1^T x_i(k) - 1^T d$ and eliminate it by adjusting the state with $y_i(k)$ to ensure the convergence point is feasible. By adopting two variables feedback to lead the state to the point satisfying the Conditions (i) and (ii) in Proposition 1, the algorithm can converge to optimal solution even with infeasible initialization and disturbance on states, which will be shown in Section V.

It is worth noting that we adopt two different constant stepizes, $\alpha$ and $\beta$ in [14] and [16], instead of decaying step-size that used in [13] and [14]. In [13], the stepsize $\alpha(k)$ is required to satisfy $\sum_{k=1}^{\infty} \alpha(k) = \infty$ and $\sum_{k=1}^{\infty} (\alpha(k))^2 < \infty$, which results in convergence rate of $O(\ln(k)/\sqrt{k})$. Similar assumptions about stepsize are also imposed in [14]. Although the algorithm proposed in [14] converges linearly, it is only suitable for fixed networks and there is always a steady state error between the convergence point and the optimal solution. In next section, we will show that the proposed algorithm converges to the optimal solution exactly at a linear convergence rate with constant stepsize even over stochastic networks.

IV. CONVERGENCE ANALYSIS OVER STOCHASTIC NETWORK

In this section, we will show that the proposed Algorithm 1 converges linearly to the optimal solution in mean square over stochastic communication networks where each communication link fails randomly. Without loss of generality, we consider the situation when $u = 1$. Extension to the case of $u > 1$ is straightforward.

Lemma 1. Let $\Gamma = \text{diag}\{\gamma_1, ..., \gamma_n\}$ and $T = L\Gamma(I - A)$, where $\eta_i \leq \gamma_i \leq \varphi_i, \forall i$. Suppose Assumptions 1-2 hold and $A$ is doubly stochastic. For any vector $v$ such that $1^T v = 0$ and $v \neq 0$, we have

$$\frac{\eta (1 - \lambda_2(A))}{\sqrt{n}} \leq \frac{\|Tv\|_2}{\|v\|_2} \leq \frac{\eta (1 - \lambda_n(A))}{\sqrt{n}}$$

where $\eta = \min_i \{\gamma_i\}$, $\varphi = \max_i \{\varphi_i\}$.

Lemma 2. Let $K_1 = \frac{\eta (1 - \lambda_2(A))}{\sqrt{n}}$ and $K_2 = \frac{\eta (1 - \lambda_n(A))}{\sqrt{n}}$.

Lemma 3. Suppose Assumption 3 holds, we have

$$\rho \left( \mathbb{E} \left\{ (W(k))^2 \right\} - \frac{11\gamma^2}{n} \right) < 1.$$

Proofs of Lemmas 1, 2 and 3 are shown in Appendices A, B and C, respectively.

Theorem 1. Let $\bar{x}_{2k} = \max_k \{\lambda_2(\mathbb{E} \{W(k)^2\})\}$, $s_1 = \alpha^2 + 2\alpha + \lambda_2$, and $s_2 = \frac{1 + \alpha^2}{\lambda_2} - 2\lambda_1 K_1$. Consider the sequence $\{x(k)\}_{k=0}^{\infty}$ and $\{y(k)\}_{k=0}^{\infty}$ generated by Algorithm 1. Suppose Assumptions 1-3 hold. For any $x(0) \in \mathbb{R}^n$ and $(\alpha, \beta)$ satisfying

$$\alpha < \frac{s_1 - \bar{x}_{2k} - 1}{\beta^2 - \frac{K_2}{K_1}}$$

there exist $m_1, m_2 > 0$ and $q_s \in (0, 1)$ such that for any $k \geq 0$,

$$\|y(k)\|_E \leq m_1 q_s^k$$

where $\nabla f_i(x_i(k)) = L \nabla f_i(x_i(k))$.

Proof. First, we multiply both sides of (15) and (16) by $\frac{11\gamma^2}{n}$, which yields

$$\bar{x}(k + 1) = \bar{x}(k) - \alpha \bar{y}(k),$$

$$\bar{y}(k + 1) = \bar{y}(k) + \bar{x}(k + 1) - \bar{x}(k).$$

Due to $y(0) = x(0) - d$, we have $y(k) = x(k) - d, \forall k$. Based on (22), we have $\bar{x}(k + 1) - d = (1 - \alpha)(\bar{x}(k) - d)$. Based on (22), we have $\bar{x}(k + 1) - d = (1 - \alpha)(\bar{x}(k) - d)$. Based on (22), we have $\bar{x}(k + 1) - d = (1 - \alpha)(\bar{x}(k) - d)$.

We can see that for any $m_0, q_s \in [1 - \alpha, 1]$, such that

$$\|y(k)\|_E \leq m_1 q_s^k$$

which means that $\bar{x}(k)$ converges in mean square linearly to $d$ and $\bar{y}(k) = \bar{x}(k) - d$ converges to 0 as long as $0 < \alpha < 2$. Due to $\bar{y}(k) = y(k) + \bar{y}(k)$, we only need to prove that there exist $m_1^*, m_2^* > 0$ and $q_s^* \in (0, 1)$ such that

$$\|\bar{y}(k)\|_E \leq m_1 q_s^k$$

from Assumptions 1 and 2, we obtain that for any $k$, there exists a diagonal matrix $\Theta(k) = \text{diag}\{\theta_1(k), ..., \theta_n(k)\}$, $\eta_i \leq \gamma_i \leq \varphi_i, \forall i$. Suppose Assumptions 1-2 hold and $A$ is doubly stochastic. For any vector $v$ such that $1^T v = 0$ and $v \neq 0$, we have

$$\frac{\eta (1 - \lambda_2(A))}{\sqrt{n}} \leq \frac{\|Tv\|_2}{\|v\|_2} \leq \frac{\eta (1 - \lambda_n(A))}{\sqrt{n}}$$

where $\eta = \min_i \{\gamma_i\}$, $\varphi = \max_i \{\varphi_i\}$.
\( \theta_i(k) \leq \varphi, \forall i, k, \) such that
\[
\nabla f(x(k + 1)) - \nabla f(x(k)) = \Theta(k)(x(k + 1) - x(k)).
\] (27)

Substituting (27) into (15), we have
\[
\nabla f(x(k + 1)) - \nabla f(x(k)) = -\alpha \Theta(k)y(k) - \beta \Theta(k)(I - W(k)) \nabla f(x(k)).
\] (28)

Multiplying both sides of (28) by \( L \), we obtain
\[
\nabla f(x(k + 1)) - \nabla f(x(k)) = -\alpha L \Theta(k)y(k) - \beta L \Theta(k)(I - W(k)) \nabla f(x(k)),
\] (29)

where \( (I - W(k)) \nabla f(x(k)) = (I - W(k))L \nabla f(x(k)) = (I - W(k)) \nabla f(x(k)) \). From (29), we further obtain
\[
\nabla f(x(k + 1)) = (I - \beta L \Theta(k)(I - W(k))) \nabla f(x(k)) - \alpha L \Theta(k)y(k) - \alpha L \Theta(k)y(k).
\] (30)

Then, from (16), we have
\[
\tilde{y}(k + 1) = W(k) \tilde{y}(k) + \tilde{x}(k + 1) - \tilde{x}(k).
\] (31)

Multiplying both sides of (15) by \( L \), we have
\[
\tilde{x}(k + 1) = \tilde{x}(k) - \alpha \tilde{y}(k) - \beta L \Theta(k)(I - W(k)) \nabla f(x(k)).
\] (32)

Substituting (32) into (31), we have
\[
\tilde{y}(k + 1) = (W(k) - \alpha L \Theta(k)) \tilde{y}(k) - \beta L \Theta(k)(I - W(k)) \nabla f(x(k)).
\] (33)

Next, we will prove (25) and (26) by mathematical induction based on (30) and (33).

It is not difficult to prove that when \( k = 0 \), there exist \( m_{1s}, m_{2s} > 0 \) and \( q_s \in (q_{0s}, 1) \) such that (25) and (26) hold.

Then, we assume that for all \( k \leq \alpha, (25) \) and (26) hold.

When \( k = \alpha + 1 \), from (33), because \( W(k) \) is independent of \( x(k) \) and \( y(k) \), we have
\[
\|y(k + 1)\|_E \leq s_1 \|y(k)\|_E + \beta(1 - \Delta_n)\|\nabla f(x(k))\|_E
\leq [s_1m_{1s} + \beta(1 - \Delta_n)m_{2s}]q^k_r.
\] (34)

From (30) and (24), we obtain based on Lemma 3 that
\[
\|\nabla f(x(k + 1))\|_E \leq s_2\|\nabla f(x(k))\|_E
+ \alpha \nabla (\|y(k)\|_E + \|f(k)\|_E)
\leq [s_2m_{1s} + \alpha \nabla (m_{1s} + m_{0})]q^k_r.
\] (35)

Due to \( \alpha < \frac{\sqrt{2 - \lambda_2}}{\sqrt{2}} - 1, \beta < \frac{\sqrt{2 - \lambda_2}}{\sqrt{2}} \), we have \( s_1 < 1 \) and \( s_2 < 1 \). Since \( \alpha \beta \frac{1 - \Delta_n}{1 - \Delta_n} < (1 - s_1)(1 - s_2) \), there exist \( m_{1s} > 0 \) and \( m_{2s} > 0 \) such that
\[
\alpha \beta \frac{1 - \Delta_n}{1 - \Delta_n}m_{1s} + m_{0}s < \beta(1 - \Delta_n)(1 - s_2)m_{2s}
< (1 - s_1)(1 - s_2)m_{1s}.
\] (36)

Then, we have there exists \( q^*_s \in (q_{0s}, 1) \) such that
\[
[s_1m_{1s} + \beta(1 - \Delta_n)m_{2s}]q^k_r \leq m_{1s}q^k_r, \quad (37)
\]
\[
[s_2m_{1s} + \alpha \nabla (m_{1s} + m_{0})]q^k_r \leq m_{2s}q^k_r, \quad (38)
\]

Substituting (37) and (38) into (34) and (35) yields
\[
\|y(k + 1)\|_E \leq m_{1s}q^k_r, \quad (39)
\]
\[
\|\nabla f(x(k + 1))\|_E \leq m_{2s}q^k_r. \quad (40)
\]

This completes the induction.

Then, based on (24) and (25), we obtain that
\[
\|y(k)\|_E \leq \|\tilde{y}(k)\|_E \leq \|y(k)\|_E \leq \|m_{0s} + m_{1s}q^k_r, \quad (39)
\]
which validates (20). The proof is completed.

**Remark 1.** In Theorem 1, the stepsize \( \alpha \) and \( \beta \) are limited by (19). The upperbound of \( \alpha \) is related to \( \beta \) and vice versa. Since \( \lim_{\alpha \to 0} \frac{\alpha}{1 - \alpha} = 0 \) and \( \lim_{\alpha \to 0} \frac{\beta}{1 - \beta} = 2K_1 \), as long as \( \alpha \) and \( \beta \) small enough, the equation (19) hold and there exist \( \alpha \) and \( \beta \) such that the proposed algorithm converges in mean square.

It should also be noted that \( m_{1s}, m_{2s} \), and \( q_s \) are independent of \( k \). If \( m_{1s} \) is a sufficiently large number, \( m_2 \) can be selected as \( \alpha \beta(1 - \lambda_2) + (1 - s_1)(1 - s_2) \) and \( q_s \) is any number in the interval \( (q_{0s}, 1) \), where \( q_{0s} = \max\{s_1 + \beta(1 - \lambda_2)m_{2s}, s_2 + \alpha \nabla (m_{2s}, q_{0s}) \} \). \( q_{0s} < 1 \) can be proved by (19). Therefore, \( m_{1s}, m_{2s} \), and \( q_s \) are all independent of \( k \).

Another result is also applicable to the following Theorems 2 and 3.

Additionally, most existing works require the network to be connected or jointly connected, i.e., there exist an integer \( B \geq 1 \) such that \( \lambda_2(W(k)W(k + 1)...W(k + B)) < 1 \). But in stochastic networks, \( B \) may not exist. In this paper, we only need the network is connected in mean, i.e., \( \lambda_2(\mathbb{E}[W(k)]) < 1 \).

**Corollary 1.** Suppose Assumptions 1-3 hold. For any \( \alpha(0) \in \mathbb{R}^n \) and \( (\alpha, \beta) \) satisfying (19), the sequence \( \{x(k)\}_{k=0}^{\infty} \) generated by Algorithm 1 converges linearly to \( x^* \) in mean square.

The convergence rate is
\[
\max \left\{ s_1 + \frac{\alpha \nabla(1 - \lambda_n)}{1 - s_2}, s_2 + \frac{\alpha \nabla(1 - \lambda_n)}{1 - s_1}, |1 - \alpha| \right\}.
\]

**Proof.** Theorem 3 implies that \( \|y(k)\|_E \) and \( \|\nabla f(x(k))\|_E \) converge to zero linearly. From (14) and (16), we have \( 1^T y(k) = 1^T x(k) - 1^T \tilde{d} \). So the condition (i) in Proposition 1 is ensured if \( \|y(k)\|_E \) converges to 0 and so is the condition (ii) in Proposition 1 if \( \|\nabla f(x(k))\|_E \) converges to 0. Therefore, \( x(k) \) converges to the optimal solution \( x^* \). Moreover, due to the linear convergence of \( \|y(k)\|_E \) and \( \|\nabla f(x(k))\|_E \), we obtain from (15) that \( x(k + 1) - x(k) \) converges linearly to 0, which, together with the above analysis, implies \( x(k) \) converges to \( x^* \) at a linear rate.

From (37) and (38), we obtain that
\[
q_s^* \geq s_1 + \beta(1 - \lambda_n)m_{1s} < \frac{s_2m_{1s} + \alpha \nabla (m_{1s} + m_{0})}{m_{2s}}. \quad (40)
\]

From (36), we have
\[
\frac{\alpha \nabla}{1 - s_2} < \frac{m_{1s}}{m_{2s}} < \frac{1 - s_1}{\beta(1 - \lambda_n)}. \quad (41)
\]

Since that \( m_{1s} \) and \( m_{2s} \) can be any positive value as long as they satisfy (41) and \( q_s^* > |1 - \alpha| \), we can get the result.

Compared with fixed networks, the stochastic network is time-varying and, most importantly, possible to be disconnected at any time, which is different from most of existing works that require the network to be always connected or jointly connected. Moreover, we obtain from Assumption 3 that as long as \( \theta_{ij}(k) > 0 \) for each communication link \( (i, j) \), we will have the relation (7). This means that the proposed algorithm converges to the optimal solution even if each communication link works with a small positive possibility. Although small \( \theta_{ij}(k) > 0 \) does not break the convergence of
the algorithm, it affects the convergence rate, which will be
extensively discussed in Section VIII.

V. RESILIENCE PROPERTIES

A. Convergence in the presence of disturbances

In this section, we show that Algorithm 1 converges to the
optimal solution in mean square under disturbance on
states. Here we do not consider the disturbance on \( y(k) \)
because \( y(k) \) is only a message that tracks the deviation while
\( x(k) \) always denotes the running status of equipments, such
as power output of generators, which is more likely to be
disturbed by misoperation, attack, or environmental noise.

We add disturbance to states of (14)-(16) and obtain that
\[
\begin{align*}
x(k + 1) &= x(k) + \zeta(k) - \alpha y(k) - \beta (I - W(k)) \nabla f(x(k)), \\
y(k + 1) &= W(k)y(k) + x(k + 1) - x(k).
\end{align*}
\]

Random variable \( \zeta(k) = [\zeta_1(k), ..., \zeta_m(k)] \) is the disturbance
injected to the state. We give the following assumption about
the disturbance.

Assumption 4. There exist \( m_\zeta > 0 \) and \( q_\zeta \in (0, 1) \) such that
for any \( k, \| \zeta(k) \|_E \leq m_\zeta q_\zeta^k \).

The resilience property means the algorithm states is able
to maintain its convergence to the optimal solution after
the injected disturbance decays. Assumption 4 restricts the
second order moment of the disturbance to be covered by
an upper bound that vanishes exponentially. Assumption 3 is
not only suitable for those continuous disturbance converging
exponentially to 0, but it also applies to all kinds of finite-
duration disturbances, i.e., the disturbance \( \zeta(k) = 0 \) after
a finite-time steps \( k \geq K_\zeta \). Based on this assumption, we prove
the linear convergence of the expectation of the states norm
\( \| x(k) \|_E \).

Theorem 2. Consider the sequence \( \{ x(k) \}_{k=0}^{\infty} \) generated by (12) and (13). Suppose Assumptions 1-4 hold. For any \( x(0) \in \mathbb{R}^n \) and \( (\alpha, \beta) \) satisfying (19), there exist \( m_{1d}, m_{2d} > 0 \) and \( q_d \in (0, 1) \) such that for any \( k \geq 0, \)
\[
\begin{align*}
\| \hat{y}(k) \|_E &= \| x(k + 1) - d \|_E \\
&\leq (1 - \alpha)^k \| x(0) - d \|_E + m_\zeta \frac{(1 - \alpha)^k - q_\zeta^k}{1 - \alpha - q_\zeta^k}.
\end{align*}
\]

Proof. Similar to the proof of Theorem 1 we have
\[
\begin{align*}
\| \hat{y}(k) \|_E &= \| x(k + 1) - d \|_E \\
&\leq (1 - \alpha)^k \| x(0) - d \|_E + m_\zeta \frac{(1 - \alpha)^k - q_\zeta^k}{1 - \alpha - q_\zeta^k}.
\end{align*}
\]

We can see that for any \( m_{0d} > \| x(0) - d \|_E + m_\zeta \frac{1}{1 - \alpha - q_\zeta^k} \) and
\( q_{0d} \in (\max\{1 - \alpha, q_\zeta^{1/k}\}, 1) \) such that
\[
\| \hat{y}(k) \|_E \leq m_{0d} q_{0d}^{k},
\]
and we only need to prove that there exist \( m_{1d}, m_{2d} > 0 \) and
\( q_d \in (q_{0d}, 1) \) such that
\[
\| \hat{y}(k) \|_E \leq m_{1d} q_{1d}^{k},
\]
\[
\| \nabla f(x(k)) \|_E \leq m_{2d} q_{2d}^{k},
\]

Similar to (50) and (53), we still have
\[
\nabla f(x(k + 1)) = (I - \beta L\Theta(k)(I - W(k))) \nabla f(x(k)) - \alpha L\Theta(k) \hat{y}(k) - \alpha L\Theta(k) \zeta(k).
\]

Next, we will prove (47) and (48) by mathematical induction
based on (49) and (50).

It is not difficult to prove that when \( k = 0 \), there exist
\( m_{1d}, m_{2d} > 0 \) and \( q_d \in (\max\{q_{0d}, q_\zeta\}, 1) \) such that (47) and
(48) hold.

Then, we assume that for all \( k \leq \kappa, (47) \) and (48) hold.
When \( k = \kappa + 1 \), from (50), because \( W(k) \) is independent of
\( x(k) \) and \( y(k) \), we have
\[
\begin{align*}
\| \hat{y}(k + 1) \|_E &\leq s_1 \| \hat{y}(k) \|_E + \beta(1 - \Delta_n) \| \nabla f(x(k)) \|_E + m_\zeta q_\zeta^k \\
&\leq [s_1 m_{1d} + \beta(1 - \Delta_n) m_{2d} + m_\zeta] q_d^k.
\end{align*}
\]

From (49) and (56), we obtain based on Lemma 3
that
\[
\begin{align*}
\| \nabla f(x(k + 1)) \|_E &\leq s_2 \| \nabla f(x(k)) \|_E + \| \zeta(k) \|_E \\
&\leq [s_2 m_{2d} + \alpha \| \zeta(k) \|_E + \| \zeta(k) \|_E] q_d^k.
\end{align*}
\]

Similar to (50) and (53), we have there exist \( m_{1d} > 0, m_{2d} > 0 \)
and \( q_d \in (\max\{q_{0d}, q_\zeta\}, 1) \) such that
\[
\begin{align*}
[s_1 m_{1d} + \beta(1 - \Delta_n) m_{2d} + m_\zeta] q_d^{k + 1} &\leq m_{1d} q_d^{k + 1},
\end{align*}
\]
\[
\begin{align*}
[s_2 m_{2d} + \alpha \| \zeta(k) \|_E + \| \zeta(k) \|_E] q_d^{k + 1} &\leq m_{2d} q_d^{k + 1}.
\end{align*}
\]

Substituting (53) and (54) into (51) and (52) yields
\[
\begin{align*}
\| \hat{y}(k + 1) \|_E &\leq m_{1d} q_d^{k + 1}, \\
\| \nabla f(x(k + 1)) \|_E &\leq m_{2d} q_d^{k + 1}.
\end{align*}
\]

This completes the induction.

Then, based on (46) and (47), we obtain that
\[
\begin{align*}
\| \hat{y}(k) \|_E &\leq \| \hat{y}(k) \|_E + \| \hat{y}(k) \|_E \leq (m_{0d} + m_{1d}) q_d^{k},
\end{align*}
\]
which validates (44). The proof is completed.

Remark 2. In Theorem 2, we consider decaying disturbance
to validate the resilience of the proposed algorithm. If the noise
does not decay to 0 but is bounded, i.e., \( q_\zeta = 1 \), \( x(k) \) cannot
converge to the optimal solution but the gap between \( x(k) \)
and the optimal solution is bounded, which can be obtained
from Theorem 2 by setting \( q_\zeta = 1 \).

Corollary 2. Suppose Assumptions 1-4 hold. For any \( x(0) \in \mathbb{R}^n \) and \( (\alpha, \beta) \) satisfying (19), there exist \( m_{0d} \in \mathbb{R}^n \) generated by (42) and (43) converges linearly to \( x^* \) in mean square. The convergence rate is
\[
\max \{ s_1 + \frac{\alpha \| \zeta(k) \|_E}{1 - s_2}, s_2 + \frac{\alpha \| \zeta(k) \|_E}{1 - s_1} \}.
\]

Corollary 2 verifies that the proposed deviation-tracking method
is able to eliminate the deviation \( 1^T \hat{x}(k) \) \( - 1^T d \)
no matter whether it is caused by infeasible initialization or
random disturbance. The reason why it is disturbance-resilient
is that we adopt deviation feedback as shown in Fig. 1. The
closed loop makes it more stable and exact than the weighted
gradient method and dual splitting method, which adopt open
loop to deal with the deviation \( 1^T \hat{x}(k) \) \( - 1^T d \).

It should be noted that both Corollary 1 and Corollary 2
shares the upper bound of stepizes and the upper bound is
independent of the parameter of disturbance, which shows
the universal resilience property of distributed deviation-tracking
algorithm.
B. Improving convergence rate

In this section, we will analyze the optimal stepsize for fastest convergence rate over fixed communication networks. The result can also be applied to stochastic networks, which will be validated in simulations.

**Theorem 3.** Let $s'_1 = \max \{\bar{\lambda}_2 - \alpha, \alpha, \alpha - \Delta_{ne}\}$ and $s'_2 = \max \{|1 - \beta K'_1|, |\beta K'_2 - 1|\}$, where $K'_1 = \frac{\eta(1-\bar{\lambda}_2)\varphi+(n-1)\eta}{\sqrt{n(1-\bar{\lambda}_2)^2+(n-1)\eta^2}}$ and $K'_2 = \frac{\varphi(1-\Delta_{ne})}{\varphi(1-\Delta_{ne})}$. Consider the sequence $\{x(k)\}_{k=0}^{\infty}$ generated by (42) and (43). Suppose Assumptions 1-4 hold and $W(k) = W_{\alpha, \beta}$ for any $x(0) \in \mathbb{R}^n$ and $(\alpha, \beta)$ satisfying

$$\alpha < 1 - \Delta_{ne}, \quad \beta < \frac{1}{K'_2},$$

$$\alpha \beta \varphi(1-\Delta_{ne}) < (1-s'_1)(1-s'_2),$$

there exist $m_1, m_2 > 0$ and $q \in (0, 1)$ such that for any $k \geq 0$,

$$\|y(k)\|_E \leq m_1 q^k,$$

$$\|\nabla f(x(k))\|_E \leq m_2 q^k.$$

The proof is shown in Appendix D.

**Theorem 4.** Suppose Assumptions 1-3 hold. The optimal stepizes of $\alpha$ and $\beta$ are expressed as

$$\beta_{op} = \frac{2}{K'_1 + K'_2},$$

$$\alpha_{op} = \min \left\{ \frac{K'_1(1-\bar{\lambda}_2)}{K'_2}, \frac{K'_1(1 + \Delta_{ne})}{2K'_1 + K'_2}, \frac{1}{2} \left( 1 + \bar{\lambda}_2 - 2\beta_{op}K'_2 + \sqrt{4(\beta_{op}K'_2 - 1)^2 + (1 - \bar{\lambda}_2)(5-\bar{\lambda}_2)} \right) \right\}.$$  

(60)

**Proof.** Similar to Corollary 2, the convergence rate is

$$\max \left\{ \frac{s'_1}{1-s'_1}, \frac{\alpha \beta \varphi(1-\Delta_{ne})}{1-s'_1} \right\} \leq \|x(k+1) - x^*\|_E / \|x(k) - x^*\|_E = \beta_{op}^{K'_2 - 1} - 1.$$  

First, we consider the optimal value of $\beta$.

Let $g_1(\alpha, \beta) = s'_1 + \frac{\alpha \beta \varphi(1-\Delta_{ne})}{1-s'_1}$ and $g_2(\alpha, \beta) = s'_2 + \frac{\alpha \beta \varphi(1-\Delta_{ne})}{1-s'_1}$. If $\beta > \frac{2}{K'_1 + K'_2}$, we obtain $s'_2 = \beta K'_2 - 1$. Then, we have

$$g_1(\alpha, \beta) = \frac{\alpha \beta \varphi(1-\Delta_{ne})}{1-s'_1},$$

$$g_2(\alpha, \beta) = \beta K'_2 - 1 + \frac{\alpha \beta \varphi(1-\Delta_{ne})}{1-s'_1}.$$  

Both $g_1(\alpha, \beta)$ and $g_2(\alpha, \beta)$ are increasing with respect to $\beta$. Similarly, we can obtain that $g_1(\alpha, \beta)$ is independent of $\beta$ and $g_2(\alpha, \beta)$ is decreasing when $\beta \leq \frac{2}{\nu_{\alpha, \beta}}$. Therefore, we obtain the best value of $\beta_{op} = \frac{2}{K'_1 + K'_2}$.

Next, we consider the optimal value of $\alpha$ when $\beta = \beta_{op}$. If $\alpha \leq \bar{\lambda}_2 + \frac{\Delta_{ne}}{2}$, we obtain $s'_1 = \bar{\lambda}_2 - \alpha$ and

$$g_1(\alpha, \beta_{op}) = \bar{\lambda}_2 - \alpha + \frac{\alpha \varphi(1-\Delta_{ne})}{K'_1},$$

$$g_2(\alpha, \beta_{op}) = \frac{K'_2 - K'_1(\bar{\lambda}_2 - \alpha)}{K'_1 + K'_2} + \frac{2K'_1}{K'_1 + K'_2} \varphi(1-\Delta_{ne}).$$

It is obvious that $g_2(\alpha, \beta)$ is increasing with respect to $\alpha$. Due to $\varphi(1-\Delta_{ne}) > 0$, $g_1(\alpha, \beta)$ is also increasing with respect to $\alpha$.

When $\alpha > \frac{\bar{\lambda}_2 + \Delta_{ne}}{2}$, we have $s'_1 = \alpha - \Delta_{ne}$,

$$g_1(\alpha, \beta_{op}) = \alpha - \Delta_{ne} + \frac{\alpha \varphi(1-\Delta_{ne})}{K'_1},$$

$$g_2(\alpha, \beta_{op}) = \frac{K'_2 - K'_1(\alpha - \Delta_{ne})}{K'_1 + K'_2} + \frac{2K'_1}{K'_1 + K'_2} \varphi(1-\Delta_{ne}).$$

Both of $g_1(\alpha, \beta_{op})$ and $g_2(\alpha, \beta_{op})$ are also increasing with respect to $\alpha$. However, $1 - \alpha$ is decreasing when $\alpha < 1$. Then the optimal $\alpha$ should be

$$\alpha_{op} = \min \{\alpha_{op1}, \alpha_{op2}\},$$

where $\alpha_{op1}$ and $\alpha_{op2}$ satisfy

$$1 - \alpha_{op1} = g_1(\alpha_{op1}, \beta_{op}); 1 - \alpha_{op2} = g_2(\alpha_{op2}, \beta_{op}).$$

Finally, we calculate the optimal $\alpha$ and obtain (60).  

The result in Theorem 4 is obtained based on the convergence result of Theorem 3. In some cases, the result in Theorem 4 may not be the optimal stepsize for fastest convergence rate, but it is still better than most stepizes within the upperbound. The results are also suitable for stochastic networks with communication failure and disturbance, which will be shown in Section VII.

VI. CONVERGENCE WITH UNCOORDINATED STEPSIZES

In above analysis, all agents employ absolutely identical stepizes. In distributed network, realizing uniform stepizes requires consensus protocol, which is based on a reliable communication network. It is difficult to coordinate these stepizes over stochastic networks. In this section, we analyze the convergence of the proposed algorithm with uncoordinated stepsizes.

For each agent $i$, the stepsizes are denoted by $\alpha_i$ and $\beta_i$. Let $\varphi = \max_i(\alpha_i)$, $\beta = \max_i(\beta_i)$, $\alpha = \min_i(\alpha_i)$, $\beta = \min_i(\beta_i)$, $D_{\alpha} = \text{diag}(\alpha_1, ..., \alpha_n)$ and $D_{\beta} = \text{diag}(\beta_1, ..., \beta_n)$. Eqs. (15) and (16) are rewritten as

$$x(k+1) = x(k) - D_{\alpha} y(k) - D_{\beta} (I - W(k)) \nabla f(x(k)),$$

$$y(k+1) = W(k) y(k) + x(k+1) - x(k).$$

We will prove (62) and (63) converge linearly to the optimal solution in mean square with uncoordinated stepsizes.

**Lemma 4.** Let $K'_1'' = \frac{\nu(1-\bar{\lambda}_2)\varphi+(n-1)\nu}{\sqrt{n(1-\bar{\lambda}_2)^2+(n-1)\nu^2}}$, $K'_2'' = \frac{\varphi(1-\Delta_{ne})}{\varphi(1-\Delta_{ne})}$ and $T_d = L D_{\beta}(I - W)$, where $W$ is a stochastic matrix satisfying Assumption 3. For any vector $v \neq 0$ such that $1^T v = 0$, we have

$$\|v - T_d v\|_E^2 \leq (1 + K'_2'' - 2K'_1'') \|v\|_E^2.$$

Lemma 4 is an extension result of Lemma 3 with replacing diagonal matrix $\Gamma$ by another diagonal matrix $\Gamma D_{\beta}$. 
Theorem 5. Let \( s_4 = 1 - \sum_{i=1}^{n} \alpha_i, \ s_5 = \pi^2 + 2\pi + \lambda_{2x}, \) and \( s_6 = \sqrt{1 + K_{2x}^2 - 2K_{2x}'}, \) Consider the sequence \( \{x(k)\}_{k=0}^{\infty} \) and \( \{y(k)\}_{k=0}^{\infty} \) generated by (62) and (63). Suppose Assumptions 1-4 hold. For any \( x(0) \in \mathbb{R}^n \) and \((\alpha, \beta)\) satisfying
\[
\sum_{i=1}^{n} \alpha_i < 2, \ \alpha < \sqrt{2 - \lambda_{2x} - 1}, \ K_{2x}'' < 2K_{2x}', \quad (64)
\]
\[
(1 - s_4)(1 - s_5)(1 - s_6) - \beta(1 - \Delta_n)\varpi(2 - s_4 - s_5) < \alpha^2(1 - s_6) + 2\alpha^2\beta(1 - \lambda_n)\varpi, \quad (65)
\]
there exist \( m_{1u}, m_{2u}, m_{3u} > 0 \) and \( q_u \in (0, 1) \) such that for any \( k \geq 0, \)
\[
\|\nabla f(x(k))\|_E \leq m_{1u}q_u^k, \quad (66)
\]
\[
\|\tilde{g}(k)\|_E \leq m_{2u}q_u^k, \quad (67)
\]
\[
\|\tilde{g}(k)\|_E \leq m_{3u}q_u^k. \quad (68)
\]

Proof. Multiplying both sides of (62) by \( \frac{11T}{n} \), we have
\[
\bar{x}(k + 1) = \bar{x}(k) - \frac{11T}{n} D_{\alpha} \bar{y}(k) - \frac{11T}{n} D_{\beta} (I - W(k)) \nabla f(x(k)). \quad (69)
\]
Then, multiplying both sides of (62) by \( L \) yields
\[
\bar{y}(k + 1) = \bar{y}(k) - LD_{\alpha} \bar{y}(k) - LD_{\beta} (I - W(k)) \nabla f(x(k)). \quad (70)
\]
Similarly, it follows from (63) that
\[
\tilde{y}(k + 1) = \tilde{y}(k) + \bar{x}(k + 1) - \bar{x}(k), \quad (71)
\]
\[
\tilde{y}(k + 1) = W(k) \tilde{y}(k) + \bar{x}(k + 1) - \bar{x}(k). \quad (72)
\]
Substituting (69) and (70) into (71) and (72) respectively yields
\[
\tilde{y}(k + 1) = (I - \frac{11T}{n} D_{\alpha}) \tilde{y}(k) - \frac{11T}{n} D_{\beta} (I - W(k)) \nabla f(x(k)), \quad (73)
\]
\[
\tilde{y}(k + 1) = (W(k) - LD_{\alpha}) \tilde{y}(k) - LD_{\beta} (I - W(k)) \nabla f(x(k)). \quad (74)
\]
Similar to (29), we have
\[
\nabla f(x(k + 1)) = (I - L\Theta(k)D_{\beta}(I - W(k))) \nabla f(x(k)) - L\Theta(k)D_{\alpha} \tilde{y}(k) - L\Theta(k)D_{\alpha} \tilde{y}(k). \quad (75)
\]

We use induction to prove (66)-(68). When \( k = 0 \), it is not difficult to find \( m_{1u}, m_{2u}, m_{3u} > 0 \) and \( q_u \in (0, 1) \) such that (66)-(68) hold. We assume when \( k \leq \kappa \), (66)-(68) hold. When \( k = \kappa + 1 \), from (73) and (74), we have
\[
\|\tilde{y}(k + 1)\|_E \leq s_4\|\tilde{y}(k)\|_E + \alpha \|\tilde{y}(k)\|_E + \beta(1 - \Delta_n)\|\nabla f(x(k))\|_E \leq s_4m_{2u}q_u^k + \alpha m_{2u}q_u^k + \beta(1 - \Delta_n)m_{1u}q_u^k, \quad (76)
\]
\[
\|\tilde{y}(k + 1)\|_E \leq s_5\|\tilde{y}(k)\|_E + \alpha \|\tilde{y}(k)\|_E + \beta(1 - \Delta_n)\|\nabla f(x(k))\|_E \leq s_5m_{3u}q_u^k + \alpha m_{3u}q_u^k + \beta(1 - \Delta_n)m_{1u}q_u^k. \quad (77)
\]
Then, it can be obtained from (75) that
\[
\|\nabla f(x(k + 1))\|_E \leq s_6\|\nabla f(x(k))\|_E + \alpha \|\nabla f(x(k))\|_E + \beta(1 - \Delta_n)\|\nabla f(x(k))\|_E \leq s_6m_{1u}q_u^k + \alpha m_{1u}q_u^k + \beta(1 - \Delta_n)m_{1u}q_u^k. \quad (78)
\]
Due to (64), we have \( s_4, s_5 \) and \( s_6 \) are all in \((0, 1)\). Then, it follows from (65) that
\[
(1 - s_4)(1 - s_5)(1 - s_6) - \beta(1 - \Delta_n)\varpi(2 - s_4 - s_5) \leq \frac{(1 - s_4)(1 - s_5)(1 - s_6) - \beta(1 - \Delta_n)\varpi}{\alpha^2(1 - s_6) + 2\alpha^2\beta(1 - \lambda_n)\varpi}. \quad (79)
\]
Based on (79), there exist \( m_{2u}, m_{3u} > 0 \) such that
\[
(\alpha^2(1 - s_6) + 2\alpha^2\beta(1 - \lambda_n)\varpi)m_{2u} \leq ((1 - s_4)(1 - s_5) - \beta(1 - \lambda_n)\varpi)m_{2u}, \quad (80)
\]
\[
(\alpha^2(1 - s_6) + 2\alpha^2\beta(1 - \lambda_n)\varpi)m_{3u} \leq ((1 - s_4)(1 - s_5) - \beta(1 - \lambda_n)\varpi)m_{3u}, \quad (81)
\]
\[
\alpha \|\nabla f(x(k))\|_E \leq \beta(1 - \Delta_n)\varpi. \quad (82)
\]
Substituting (80)-(82) into (76)-(78) leads to (66)-(68) when \( k = \kappa + 1 \). This completes the proof. \( \square \)

From Theorem 5, we can deduce that the convergence point of \( x(k) \) satisfies the two conditions in Proposition 1, which means that the proposed algorithm still converges linearly to the optimal solution in mean square even with uncoordinated stepsizes. Thus, agents do not need to coordinate stepsizes before implementing the algorithm.

VII. NUMERICAL RESULTS

In this section, simulations are provided to validate the effectiveness of the proposed deviation-tracking algorithm (DTA). Considering a network of 10 agents, of which the topology is randomly generated. The cost functions are chosen to be quadratic functions:
\[
f_i(x) = a_ix_i^2 + b_ix_i + c_i, \ \forall i, \quad (75)
\]
which are widely used in economic dispatch problem to model the cost induced by producing certain amount of power \([25]\). The values of \( a, b \) and \( c \) are randomly chosen within the ranges that are used in \([13]\). Their values are shown as follows.

| Agent | 1 | 2 | 3 | 4 | 5 |
|-------|---|---|---|---|---|
| \( a_i \) | 0.0314 | 0.1242 | 0.0892 | 0.0379 | 0.0366 |
| \( b_i \) | 0.352 | 0.349 | 0.278 | 0.331 | 0.234 |
| \( c_i \) | 0 | 0 | 0 | 0 | 0 |

A. Convergence of deviation-tracking algorithm

First, we simulate the convergence results of deviation-tracking algorithm with communication failure while no disturbance is injected. Each communication link fails with
probability $1 - \theta$, where $\theta = 0.5$. We select the stepsize $\alpha$ to be $0.1\alpha_{op}, \ 0.3\alpha_{op}, \ 0.5\alpha_{op}, \ 0.7\alpha_{op}, \ 0.8\alpha_{op}, \ 0.9\alpha_{op}, \ 0.95\alpha_{op}$ and $\alpha_{op}$, where $\alpha_{op}$ is calculated by (60). The stepsizes are all constant and satisfy (19). We also show the convergence result when $\alpha$ equals the upper bound, which is obtained by (19). Another stepsize $\beta$ is fixed to be $\beta_{op}$. We use $\|x(k) - x^\star\|_2$ to describe the distance between the state at time $k$ and the optimal solution. The results are shown in Fig. 2.

We obtain from Fig. 2 that the state converge to the optimal solution with all the selected constant stepsizes. This confirms Theorem 1 and suggests that the proposed algorithm does not need decaying stepsize to ensure convergence. The advantage of adopting constant stepsizes is also shown in Fig. 2 that $x(k)$ converges at a linear rate for all the chosen stepsizes. Moreover, it is worth noting that when $\alpha \leq \alpha_{op}$, the convergence rate become faster when $\alpha$ increases and when $\alpha = \alpha_{op}$, the algorithm converges faster than that with other selected stepsizes except the upper bound. In fact, the optimal stepsize should be extremely close to the upper bound. The reason why there exists a small gap between them is that the theoretical result is obtained based on generic functions defined in Assumptions 1 and 2 instead of specific quadratic functions. It is, however, observed that the theoretical result is quite close to the optimal stepsize, which suggests that the result of Theorem 4 is also suitable for stochastic networks.

To validate the optimal stepsize of $\beta$, i.e. $\beta_{op}$, calculated in Theorem 4 we select 14 different values of $\beta$, i.e., $0.1\beta_{op}, \ 0.3\beta_{op}, \ 0.5\beta_{op}, \ 0.7\beta_{op}, \ 0.8\beta_{op}, \ 0.9\beta_{op}, \ \beta_{op}, \ 1.01\beta_{op}, \ 1.02\beta_{op}, \ 1.04\beta_{op}, \ 1.06\beta_{op}, \ 1.07\beta_{op}, \ 1.075\beta_{op}, \ 1.08\beta_{op}$. The stepsize $\alpha$ is fixed as $\alpha_{op}$. The convergence rate is calculated by

$$q_n = \frac{\|x(k) - x_{op}\|_2}{\|x(k-1) - x_{op}\|_2},$$

where $k_e$ denotes the index of the last iteration and $N$ means the amount of the sample. In this experiment, we choose $k_e = 25000$ and $N = 1000$. The result is shown in Fig. 3. We find that the convergence rate $q_n$ decrease gradually first and then increase rapidly. The numerical optimal $\beta$ is quite close to the theoretical result $\beta_{op}$.

To validate the convergence with uncoordinated stepsizes, we randomly generate stepsizes within the range defined by (64) and (65). The maximum and minimum of these stepsizes are denoted by $\bar{\theta}$ and $\underline{\theta}$, respectively. We first simulate the convergence with these uncoordinated stepsizes. Then, we plot the convergence results when $\alpha_i = \bar{\theta}, \forall i$ and $\alpha_i = \underline{\theta}, \forall i$, respectively. The convergence results are shown in Figure 4. The figure suggests that the proposed algorithm converges linearly to the optimal solution with uncoordinated stepsizes.

We find that the states converge at a faster rate than that when the all agents’ stepsizes equal to $\alpha_{\text{opt}}$, and slower than that when $\alpha_i = \bar{\theta}$. So, if $\bar{\theta} - \underline{\theta}$ is small, we can obtain an estimate of the convergence rate with uncoordinated stepsizes.

Next, we analyze the convergence results with disturbance injected in states. The disturbances we choose to inject to the states are Gaussian. To meet the requirement in Assumption 4 we set the variance of $\zeta_i(k), \forall i$ converges to 0 linearly. We select the same eight values as above for stepsize $\alpha$ and set $\beta = \beta_{op}$. The results are shown in Fig. 5. We can see that states still converge to the optimal solutions even with disturbance. With all the selected stepsizes, the algorithm still converges at a linear rate. Moreover, we find that when $\alpha \leq 0.7\alpha_{op}$, the convergence rate is faster as $\alpha$ grows. But $\alpha > 0.7\alpha_{op}$, the convergence rate does not change obvious and greatly affected by the disturbance. This is because the convergence rate is also influenced by the decaying rate of the disturbance, i.e. $q_{\zeta}$ in (61), which is independent of $\alpha$.

To find how the probability of failure influence convergence rate. We select 10 different value of $\theta$, i.e., 0.9, 0.1, 0.05, 0.03, 0.025, 0.02, 0.015, 0.01, 0.05 and 0. The results are shown Fig. 6. We can see that when $\theta > 0.05$, the conver-
gence rate does not change. The reason is that $|1 - \alpha| > \max \{s'_1 + \alpha s_2(1 - \frac{1}{s'_1}), s'_2 + \alpha s_3(1 - \frac{1}{s'_2})\}$ and $q' = |1 - \alpha|$, where $|1 - \alpha|$ is independent of $\theta$ while $s'_2$ increases as $\theta$ grows. When $\theta \leq 0.03$, we find the convergence rate reduces when $\theta$ increases. Especially, when $\theta = 0$ the algorithm cannot converge to the optimal solution because the network becomes a fixed disconnected network, which does not satisfy the requirement of Assumption 3.

B. Comparison with the state-of-art

Finally, we compare the proposed algorithm with weighted gradients algorithm (WGA) [9] and dual splitting algorithm (DuSPA) [15]. We consider two types of disturbance: Gaussian disturbance and Laplace disturbance. To meet the requirement of Assumption 4, we set the mean and the variance of disturbance according to Assumption 3. We add these kinds of disturbance to DTA, WGA and DuSPA and obtain the results shown in Fig. 7. We can see that utilizing the deviation-tracking method, DTA still converges linearly to the optimal solution even under disturbance while WGA and DuSPA cannot converge to the optimal solution. To find why WGA cannot converge to the optimal solution, we obtain the following proposition.

**Proposition 2.** Considering the following WGA,

$$x_i(k + 1) = x_i(k) + \zeta_i(k) - \alpha \sum_{j=1}^{n} w_{ij}(\nabla f_i(x_i(k)) - \nabla f_j(x_j(k))), \quad (83)$$

where the initial value $x_i(0), \forall i$ satisfies $\sum_{i=1}^{n} x_i(0) = \sum_{i=1}^{n} d_i$, and $\zeta_i(k)$ is the disturbance added to $x_i(k)$. $W = \{w_{ij}\}_{n \times n}$ is doubly stochastic. Then, at least one of the following does not hold.

(i) $x_i(k), \forall i$ converges to a feasible $x_i(\infty), \forall i$, which satisfies $\sum_{i=1}^{n} x_i(\infty) = \sum_{i=1}^{n} d_i$;

(ii) $\sum_{k=0}^{\infty} \sum_{i=1}^{n} \zeta_i(k) \neq 0$.

The proof of Proposition 2 is shown in Appendix E. Proposition 2 shows that $\sum_{k=0}^{\infty} \sum_{i=1}^{n} \zeta_i(k) = 0$ is a necessary condition for (i). This implies that the effect from $\zeta_i(k)$ is accumulative and the perturbation of even only one agent at one iteration causes a deviation of convergence. This is the reason why WGA cannot converge to the optimal solution under disturbance. This Proposition is also applicable to DuSPA because $y(k)$ in [15] is also required to be feasible through all iterations. Any disturbance injected to $y(k)$ leads to infeasible states. Figure 7 shows that the gaps between states and the optimal solutions caused by disturbance cannot be eliminated by WGA and DuSPA, which validates Proposition 2. Compared with WGA and DuSPA, DTA is disturbance resilient as shown in Theorem 2.

VIII. Conclusion

In this paper, we studied the distributed resource allocation problem over stochastic networks subject to random communication failure and disturbance on state. Based on the conditions of optimal solution, we proposed a disturbance-resilient distributed resource allocation algorithm by using deviation-tracking methods. Different from most existing algorithms that are only suitable for fixed networks, the proposed algorithm applies to stochastic networks. The algorithm was proved to converge linearly to the optimal solution in mean square even with communication failure. Moreover, we find that the proposed algorithm has resiliency to disturbance, i.e., our algorithm still converges to the optimal solution in mean square even under the disturbance on states. Based on the convergence result, we provided a method to improve the convergence rate. It was also proved that the proposed algorithm converges to the optimal solution in mean square even with uncoordinated stepsizes. Future works will focus on constrained distributed resource allocation and application of the algorithm to smart grid.

APPENDIX A

**PROOF OF LEMMA 1**

By definition, we obtain

$$\frac{\|Tv\|_2}{\|v\|_2} \leq \rho(T) \leq \rho(L) \rho(\Gamma) \rho(I - A) \leq \varphi(1 - \lambda_n(A)).$$

This proves the right most inequality.

Next, consider the left inequality of (17). Due to Assumption 3, we have $1^T(I - A)v = 0$. Let $z = (I - A)v$, $z = [z_1, z_2, ..., z_n]^T$, $z_i \in \mathbb{R}, \forall i$. For that $1^Tv = 0$ and $v \neq 0$, we have $1^Tz = 0$ and $z \neq 0$, so there exists a full permutation of $\{1, ..., n\}$, which is denoted by $t_1, ..., t_n$, and $t \in \{1, ..., n - 1\}$ such that $z_{t_i} \geq 0, \forall i \in \{0, ..., t\}$, $z_{t_i} < 0, \forall i \in \{t + 1, ..., n\}$ and $\sum_{i=1}^{t} z_{t_i} = -\sum_{i=t+1}^{n} z_{t_i}$.

Since $L = I - \frac{11^T}{n}$ and $\frac{11^T}{n}$ has only one nonzero eigenvalue, whose corresponding eigenvectors lie in span{1}, we consider the projection of $\Gamma(I - A)v$ on span{1}, i.e.,
\[ \frac{\|1 - \frac{\lambda}{\hat{\Gamma}}(I - A)v \|_2^2}{\|1 - \lambda/\Gamma(I - A)v \|_2^2} = \frac{1}{\lambda} \sum_{i=1}^n \gamma_i z_i, \text{ Then,} \]
\[ \frac{\|1 - \frac{\lambda}{\hat{\Gamma}}(I - A)v \|_2^2}{\|1 - \lambda/\Gamma(I - A)v \|_2^2} \leq \frac{\left( \sum_{i=1}^n \gamma_i z_i \right)^2}{\left( \sum_{i=1}^n \gamma_i z_i \right)^2} = (n - 1) \left( \gamma + \frac{n}{\eta} \right)^2 \leq \frac{\left( \gamma - \frac{n}{\eta} \right)^2}{\left( \gamma + \frac{n}{\eta} \right)^2} \leq \frac{(n - 1) \left( \gamma - \frac{n}{\eta} \right)^2}{\left( \gamma + \frac{n}{\eta} \right)^2}. \]

It thus follows that \[ \frac{\|1 - \frac{\lambda}{\hat{\Gamma}}(I - A)v \|_2^2}{\|1 - \lambda/\Gamma(I - A)v \|_2^2} \geq \frac{\left( \gamma + \frac{n}{\eta} \right)^2}{\left( \gamma - \frac{n}{\eta} \right)^2}. \]

This completes the proof of (7).

### Appendix B
#### Proof of Lemma 2

We decompose \( (I - \beta T_s) v \) and obtain
\[ E\{\| (I - \beta T_s) v \|_2^2 \} = E\{v^T (I - \beta T_s)^T (I - \beta T_s) v \} = E\{v^T (I + \beta^2 T_s^2 - \beta T_s^2 - \beta T_s) v \} = E\{v^T (I + \beta^2 \{T_s^2 v \}) - 2E\{v^T T_s v \} \}. \]

First, we consider \( E\{T_s v \|^2 \}. \) Since
\[ \frac{\|T_s v \|^2}{\|v \|^2} \leq \rho(T_s) \leq \rho(L) \rho(T) \rho(I - W) \leq \gamma(1 - \Delta_n), \]

We have \( E\{T_s v \|^2 \} \leq \gamma^2 (1 - \Delta_n) \rho^2 \{v \|^2 \}. \)

Next, we will find the lower bound of \( E\{v^T T_s v \}. \)

Similar to (85), we have
\[ \|L(I - E\{W\})v\|_2 \geq \frac{n^2(1 - \lambda_2^2) \left( \gamma + (n - 1)\eta \right)^2}{n \left( \gamma^2 + (n - 1)\eta^2 \right)} \]

Therefore, for any vector \( v \) such that \( T^v v = 0 \) and \( \gamma \neq 0, \)
\[ K_1 \leq \|L(I - E\{W\})v\|_2 \leq K_2. \]

Because \( L \) and \( I - E\{W\} \) are positive semi-definite, \( \Gamma \) is positive definite, we get that the eigenvalues of \( L(I - E\{W\}) \) are nonnegative. Then, together with (85), we have for any sufficiently small \( \gamma > 0, \)
\[ \| (I - \gamma L(I - E\{W\})) v \|_2 \leq (1 - \gamma K_1) \| v \|_2. \]

On the other hand,
\[ \| (I - \gamma L(I - E\{W\})) v \|_2 = \| v \|^2 + \gamma^2 \| L(I - E\{W\}) v \|_2 \]
\[ \geq (1 + \gamma^2 K_1) \| v \|^2 - 2\gamma^2 \| L(I - E\{W\}) v \|_2. \]

Combining (89) with (90), we have \( (1 + \gamma^2 K_1) \| v \|^2 - 2\gamma^2 \| L(I - E\{W\}) v \|_2 \leq (1 - \gamma K_1) \| v \|^2 \). Then, we obtain \( \gamma^2 \| L(I - E\{W\}) v \|_2 \geq K_1 \| v \|^2. \) Since \( L \) and \( \Gamma \) is constant, we have \( E\{v^T L(I - E\{W\}) v \} \geq K_1 E\{\| v \|^2 \}. \)

Substituting it into (86), we have
\[ E\{\| (I - \beta T_s) v \|^2 \} \leq (1 + \beta^2 K_2^2 - 2\beta K_1) E\{\| v \|^2 \}. \]

This completes the proof.

### Appendix C
#### Proof of Lemma 3

Let \( E_1(k) = \{ (i,j) | \{ P \{ w_{ij}(k) > 0 \} > 0 \} \} \) and \( E_2(k) = \{ (i,j) | \{ P \{ w_{ij}(k) > 0 \} > 0 \} \}, \) where \( w_{ij}(k) = \sum_{i=1}^n w_{ij}(k) = \sum_{i=1}^n w_{it}(k) w_{ij}(k). \) Since \( P \{ w_{ij}(k) > 0 \} = 1, \forall i, j \) and \( P \{ w_{it}(k) > 0 \} = 1, \forall i, \) we have that \( \forall i, j, \)
\[ P \{ w_{ij}(k) > 0 \} \geq P \{ w_{ij}(k) w_{ij}(k) > 0 \} = P \{ w_{ij}(k) > 0 \}, \]
which suggests \( E_1(k) \subseteq E_2(k). \) From Assumption 3, \( E_1(k) \) is connected, so, \( E_2(k) \) is also connected. Then, we can obtain (17). This completes the proof.

### Appendix D
#### Proof of Theorem 3

Similar to the proof of Theorem 1, it is not difficult to obtain that there exist \( m_0 > 0 \) and \( \max (1 - \alpha, \gamma) < \gamma_0 < 1 \) such that \( \|y(k)\|^2 \leq w_0 q_0. \) We can also obtain that
\[ \nabla f(x(k + 1)) = (I - \beta L\Theta(k) (I - W)) \nabla f(x(k)) - \alpha L\Theta(k) y(k) - \alpha L\Theta(k) y(k) + L\Theta(k) c(k). \]

Next, we will prove \( \|\gamma(k)\| \leq m_1^2 \gamma^2 \) and \( \|\nabla f(x(k))\| \leq m_2^2 \gamma^2 \) by mathematical induction based on (92) and (93).

It is not difficult to prove that when \( k = 0, \) there exist \( m_1, m_2 > 0 \) and \( \gamma' \in (q_0, 1) \) such that (25) and (26) hold.

Then, we assume that for all \( k \leq \gamma, (25) \) and (26) hold. When \( k = \gamma + 1, \) from (23), we have
\[ \|\gamma(k + 1)\| \leq \|s_1 m_1 \beta (1 - \Delta_n) m_2 + m_3 \| q^k. \]

From (30) and (24), we obtain
\[ \|\nabla f(x(k + 1))\| \leq \|s_2 m_2 + \alpha \gamma (m_1 + m_0) \| m_3 \| q^k. \]

Due to \( \alpha < 1 - \Delta_n, \) \( \beta < \frac{2}{K_2}, \) we have \( s_1 < \) and \( s_2 < \). Since \( \alpha \beta \gamma (1 - \Delta_n) < (1 - s_1)(1 - s_2), \) similar to (37) and
We prove that if (ii) holds, (i) does not hold. Since $W$ is doubly stochastic, we have \[
\sum_{i=1}^{n} x_i(k) = \sum_{i=1}^{n} \sum_{j=1}^{m} \delta_{ij} = \sum_{i=1}^{n} \sum_{j=1}^{m} \zeta_{ij}(k).
\]
Then, we get \[
\sum_{i=1}^{n} x_i(k) - \sum_{i=1}^{n} \sum_{j=1}^{m} \zeta_{ij}(k) = 0.
\]
We can see that if (ii) holds, \[
\sum_{i=1}^{n} x_i(\infty) \neq \sum_{i=1}^{n} \sum_{j=1}^{m} \xi_{ij},
\]
so (i) does not hold. This completes the proof.

**Appendix E**

**Proof of Proposition 2**

We prove that if (ii) holds, (i) does not hold. Since $W$ is doubly stochastic, we have \[
\sum_{i=1}^{n} x_i(k) = \sum_{i=1}^{n} \sum_{j=1}^{m} \delta_{ij} = \sum_{i=1}^{n} \sum_{j=1}^{m} \zeta_{ij}(k).
\]
Then, we get \[
\sum_{i=1}^{n} x_i(k) = \sum_{i=1}^{n} \sum_{j=1}^{m} \zeta_{ij}(k).
\]
We can see that if (ii) holds, \[
\sum_{i=1}^{n} x_i(\infty) \neq \sum_{i=1}^{n} \sum_{j=1}^{m} \xi_{ij},
\]
so (i) does not hold. This completes the proof.

**REFERENCES**

[1] J. Qin, Y. Wan, X. Yu, F. Li, and C. Li, “Consensus-based distributed coordination between economic dispatch and demand response,” *IEEE Transactions on Smart Grid*, vol. 10, no. 4, pp. 3709–3719, 2019.

[2] 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.

[3] X. He, J. Yu, T. Huang, and C. Li, “Distributed power management for dynamic economic dispatch in the multicrogirds environment,” *IEEE Transactions on Control Systems Technology*, vol. 27, no. 4, pp. 1651–1658, 2019.

[4] H. Zhao, K. Ding, N. I. Sarkar, J. Wei, and J. Xiong, “A simple distributed channel allocation algorithm for d2d communication pairs,” *IEEE Transactions on Vehicular Technology*, vol. 67, no. 11, pp. 10960–10969, 2018.

[5] F. Deng, J. Luo, L. He, Q. Liu, X. Li, and L. Cai, “Cooperative channel allocation and scheduling in multi-interface wireless mesh networks,” *Peer-to-Peer Networking and Applications*, vol. 12, no. 1, pp. 1–12, 2019.

[6] T. X. Tran and D. Pompili, “Joint task offloading and resource allocation for multi-server mobile-edge computing networks,” *IEEE Transactions on Vehicular Technology*, vol. 68, no. 1, pp. 856–868, 2019.

[7] M. B. Qureshi, M. M. Dehnavi, N. Min-Allah, M. S. Qureshi, H. Husain, I. Rentfous, N. Tziritas, L. Loukopoulos, S. U. Khan, C. Xu, and A. Y. Zomaya, “Survey on grid resource allocation mechanisms,” *Journal of Grid Computing*, vol. 12, no. 2, pp. 399–441, 2014.

[8] J. Xu, S. Zhu, Y. C. Soh, and L. Xie, “A bregman splitting scheme for distributed optimization over networks,” *IEEE Transactions on Automatic Control*, vol. 63, no. 11, pp. 3809–3824, 2018.

[9] H. Lakshmanan and D. P. De Farias, “Decentralized resource allocation in dynamic networks of agents,” *SIAM Journal on Optimization*, vol. 19, no. 2, pp. 911–940, 2008.

[10] I. Necoara, “Random coordinate descent algorithms for multi-agent convex optimization over networks,” *IEEE Transactions on Automatic Control*, vol. 58, no. 8, pp. 2001–2012, 2013.

[11] T. T. Doan and A. Olshovsky, “Distributed resource allocation on dynamic networks in quadratic time,” *Systems & Control Letters*, vol. 99, pp. 57–63, 2016.

[12] P. Yu, Y. Hong, and F. Liu, “Initialization-free distributed algorithms for optimal resource allocation with feasibility constraints and application to economic dispatch of power systems,” *Automatica*, vol. 74, no. 1, pp. 259–269, 2016.

[13] T. T. Doan and C. L. Beck, “Distributed resource allocation over dynamic networks with uncertainty,” *arXiv preprint arXiv: 1708.03543v9*, 2018.

[14] S. Kar and G. Hug, “Distributed state estimation and energy management in smart grids: A consensus+ innovations approach,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 6, pp. 1022–1038, 2014.

[15] J. Xu, S. Zhu, Y. C. Soh, and L. Xie, “A dual splitting approach for distributed resource allocation with regularization,” *IEEE Transactions on Control of Network Systems*, vol. 6, no. 1, pp. 403–414, 2019.

[16] M. Zhlobbarysov, C. N. Hadjicostis, and A. D. Dominguez-Garcia, “Fast distributed coordination of distributed energy resources over time-varying communication networks,” *arXiv preprint arXiv:1907.07604v4*, 2020.

[17] W. Du, L. Yao, D. Wu, X. Li, G. Liu, and T. Yang, “Accelerated distributed energy management for microgrids,” *2018 IEEE Power & Energy Society General Meeting*, pp. 1–5, 2018.

[18] Z. Zhang and M. Y. Chow, “Convergence analysis of the incremental cost consensus algorithm under different communication network topologies in a smart grid,” *IEEE Transactions on Power Systems*, vol. 27, no. 4, pp. 1761–1768, 2012.

[19] S. Kar, G. Hug, J. Mohammadi, and J. M. F. Moura, “Distributed state estimation and energy management in smart grids: A consensus+ innovations approach,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 6, pp. 1022–1038, 2014.

[20] S. Yang, S. Tan, and J. X. Xu, “Consensus based approach for economic dispatch problem in a smart grid,” *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 4412–4416, 2013.

[21] H. F. Habib, C. R. Lashway, and O. A. Mohammed, “A review of communication failure impacts on adaptive microgrid protection schemes and the use of energy storage as a contingency,” *IEEE Transactions on Industry Applications*, vol. 54, no. 2, pp. 1194–1207, 2018.

[22] J. Xu, S. Zhu, Y. C. Soh, and L. Xie, “Convergence of asynchronous distributed gradient methods over stochastic networks,” *IEEE Transactions on Automatic Control*, vol. 63, pp. 434–448, 2018.

[23] S. Kar, J. M. F. Moura, and K. Ramanan, “Distributed parameter estimation in sensor networks: Nonlinear observation models and imperfect communication,” *IEEE Transactions on Information Theory*, vol. 58, no. 7, pp. 3575 – 3605, 2012.

[24] G. Gordon and R. Tibshirani, “Karush-Kuhn-Tucker conditions,” *Optimization*, vol. 725, no. 1, pp. 10–36, 2012.

[25] A. J. Wood, B. F. Wollenberg, and G. B. Sheblé, *Power generation, operation, and control*, John Wiley & Sons, 2013.
Cailian Chen (S’03-M’06) received the B.Eng. and M.Eng. degrees in automatic control from Yanshan University, Qinhuangdao, China, in 2000 and 2002, respectively, and the Ph.D. degree in control and systems from the City University of Hong Kong, Hong Kong, in 2006.

She was a Senior Research Associate with the City University of Hong Kong, in 2006 and a Post-Doctoral Research Associate with the University of Manchester, Manchester, U.K., from 2006 to 2008. She joined the Department of Automation, Shanghai Jiao Tong University, Shanghai, China, in 2008 as an Associate Professor, where she is currently a Full Professor. She was a Visiting Professor with the University of Waterloo, Waterloo, ON, Canada, from 2013 to 2014. She has authored or co-authored two research monographs and over 100 referred international journal and conference papers. She has invented over 20 patents. Her current research interests include wireless sensor networks and industrial applications, computational intelligence and distributed situation awareness, cognitive radio networks and system design, Internet of Vehicles and applications in intelligent transportation, and distributed optimization.

Dr. Chen was a recipient of the prestigious IEEE Transactions on Fuzzy Systems Outstanding Paper Award in 2008 and the Best Paper Award of the Ninth International Conference on Wireless Communications and Signal Processing in 2017. She was the recipient of the First Prize of Natural Science Award twice from the Ministry of Education of China in 2006 and 2016, respectively. She was honored as the “Changjiang Young Scholar” by the Ministry of Education of China in 2015 and “Excellent Young Researcher” by the NSF of China in 2016.

Xinping Guan (M’04-SM’04-F’18) is currently a Chair Professor with Shanghai Jiao Tong University, Shanghai, China, where he is the Dean of the School of Electronic, Information and Electrical Engineering, and the Director of the Key Laboratory of Systems Control and Information Processing, Ministry of Education of China. He has authored or co-authored 4 research monographs, over 270 papers in IEEE TRANSACTIONS and other peer-reviewed journals and numerous conference papers. His current research interests include industrial cyber-physical systems, wireless networking and applications in smart city and smart factory, and underwater sensor networks.

Dr. Guan was a recipient of the Second Prize of the National Natural Science Award of China in 2008, the First Prize of the Natural Science Award from the Ministry of Education of China in both 2006 and 2016, the First Prize of the Technological Invention Award of Shanghai Municipal, China, in 2017, the IEEE Transactions on Fuzzy Systems Outstanding Paper Award in 2008, the National Outstanding Youth Honored by the NSF of China, the Changjiang Scholar by the Ministry of Education of China, and the State-Level Scholar of New Century Bai Qianwan Talent Program of China. As a Principal Investigator, he has finished/been working on many national key projects. He is the Leader of the prestigious Innovative Research Team of the National Natural Science Foundation of China. He is an Executive Committee member of the Chinese Automation Association Council and the Chinese Artificial Intelligence Association Council.