A dual splitting approach for distributed resource allocation with regularization

Xu, Jinming; Zhu, Shanying; Soh, Yeng Chai; Xie, Lihua

2018

Xu, J., Zhu, S., Soh, Y. C., & Xie, L. (2019). A dual splitting approach for distributed resource allocation with regularization. *IEEE Transactions on Control of Network Systems*, 6(1), 403-414. doi:10.1109/TCNS.2018.2834310

<https://hdl.handle.net/10356/142714>

<https://doi.org/10.1109/TCNS.2018.2834310>

© 2018 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. The published version is available at: <https://doi.org/10.1109/TCNS.2018.2834310>.

Downloaded on 26 Apr 2023 08:00:49 SGT

A Dual Splitting Approach for Distributed Resource Allocation with Regularization

Jinming Xu, Shanying Zhu, *Member, IEEE*, Yeng Chai Soh, and Lihua Xie, *Fellow, IEEE*

Abstract—We deal with a class of distributed resource allocation problems where each agent attempts to minimize its own cost while respecting network-wide resource constraints as well as local capacity limits. This problem arises from many areas such as economic dispatch, network utility maximization and demand response. Most existing methods are centralized while few works are devoted to solving the problem in a distributed manner. The problem becomes even more challenging when there is a (non-smooth) regularization term in the cost function. In this paper, we propose a novel distributed algorithm (termed DuSPA) to solve the above problem based on duality analysis and splitting methods. For privacy concerns, this algorithm is not required to communicate sensitive gradient information while still achieving the optimum without sacrificing the performance. We will show that the proposed algorithm converges at a *non-ergodic* convergence rate of $O(\frac{1}{k})$ for general convex cost functions and a linear convergence rate for smooth and strongly convex cost functions, respectively. Further, we apply the proposed algorithm to an economic dispatch problem to show its effectiveness.

Index Terms—Splitting methods, resource allocation, regularization, optimization duality.

I. INTRODUCTION

DISTRIBUTED resource allocation has received much attention recently due to its wide applications in various domains, such as economic dispatch in power systems [2], resource allocation in wireless networks [3] and task assignment in robotic networks [4]. Resource allocation is a process of assigning available resources to users in a way that is optimal to the society. Due to limited resources, agents have to act in a restricted way while taking into account others' interests. The conventional centralized or parallel approaches, which rely on entire network information, may not be scalable considering the large volume of agents scattered over a wide area [5]. Moreover, the operation of the whole system will

be terminated once the center taking charge of the system is non-functional either due to failures or attacks. Alternatively, a distributed approach, which works with merely local information, is more robust to possible node failures and scalable to the size of the network. Due to privacy concerns, the data used to transmit the information over the network is also a key factor to be considered in the algorithm design.

Motivated by the fact that the optimal solution for the resource allocation problem is attained when the marginal costs are all equal [6], the well-known consensus protocol [7] can be thus employed to solve the problem. In particular, Xiao and Boyd [8] studied a center-free algorithm and provided sufficient conditions on the weight matrix to guarantee its convergence. Although this algorithm is simple, it does not take into account local constraints of agents, e.g., the capacity limit of power output of the generators in a power system. To account for local constraints, instead of directly solving the above primal problem, Doan and Alex [9] attempted to tackle its dual counterpart for which most recent distributed algorithms can be applied [10]–[13]. Following this idea, they proposed a distributed primal-dual method which is able to solve the problem even under varying networks and enjoys a linear convergence rate so long as the stepsize is sufficiently small for strongly convex and smooth cost functions. A similar idea appeared in [14] where the authors proposed a general efficient algorithm which can solve the above primal problem as well as its dual simultaneously. However, when it comes to composite problems where the cost function contains a (potentially non-smooth) regularization term, the dual problem in a conjugate form will be no longer splittable [15], making it unable to be solved efficiently. Due to the fact that most utility functions are strongly concave in network utility maximization (NUM) problems, Beck *et. al* also proposed to solve it from its dual using fast gradient methods [16]. However, this method is dedicated to problems where the physical system naturally exhibits a distributed structure, e.g., NUM problems.

As a parallel line of work, in economic dispatch problems, similar algorithms have been developed to determine the optimal power output of all generators that leads to the minimum operating cost while respecting the system constraints, such as the capacity limit of the power output of generators and overall load demand. In particular, in [17] a parallel algorithm was proposed to solve the economic dispatch problem which requires a centralized processor to ensure the exact match between the overall power supply and demand. To eliminate the centralized node, a consensus + innovation algorithm was designed which only requires local mismatch information for innovation purpose [18]. Although this algorithm is completely

This work was supported in part by Singapore's National Research Foundation under NRF2011NRF-CRP001-090 and in part by the Energy Research Institute at NTU (ERI@N). The work of S. Zhu was supported in part by National Key R&D Program of China under 2016YFB0901900, NSF of China under Grant 61603251 and NSF of Shanghai under grant 18ZR1419900. The work of L. Xie was supported by National Research Foundation of Singapore under NRF-CRP 8-2011-03 and Projects of Major International (Regional) Joint Research Program NSFC of China under Grant 61720106011. This work was partially presented in the 56th CDC, Melbourne, 2017 [1].

Jinming Xu is with the Ira A. Fulton Schools of Engineering, Arizona State University, Tempe, AZ 85281 USA (e-mail: xuji0016@e.ntu.edu.sg)

Yeng Chai Soh and Lihua Xie are with School of Electrical and Electronic Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798. E-mails: xuji0016@e.ntu.edu.sg and {EYCSOH, ELHXIE}@ntu.edu.sg

Shanying Zhu is with Department of Automation, Shanghai Jiao Tong University, and Key Laboratory of System Control and Information Processing, Ministry of Education of China, Shanghai 200240, China. E-mail: shyzhu@sjtu.edu.cn

distributed, it requires the stepsize to be decaying to guarantee the exact convergence, resulting in slow convergence speed. To overcome this drawback, in [19], a novel distributed algorithm was developed by introducing an extra step of consensus for the sake of estimating the overall mismatch in a distributed way. They showed that the algorithm converges as long as certain parameter is sufficiently small though no specific upperbound is given therein. Interestingly, this algorithm turns out to be a special case of the aforementioned primal-dual method [9]. Likewise, in continuous domain, similar approaches have been developed. In particular, in [20], an exact penalty method is proposed to solve the problem based on the center-free algorithm [8]. They also developed an initialization-free algorithm using dynamic average consensus and integral feedback which essentially make it adaptive to varying loads [21]. However, these algorithms require the agents to transmit directly the gradient information over the network, which is not desirable due to privacy concerns. This drawback is overcome by the methods proposed in [22], [23] which require the cost function to be continuously differentiable even up to the second order. To deal with non-smooth cost functions, two different algorithms are developed for solving general extended monotropic problem in [24]. However, they still require the cost function to be strictly convex and no explicit rate is obtained therein.

In this paper, we consider a general *composite* resource allocation problem which subsumes many real applications. Instead of directly solving the primal problem, we attempt to solve its dual counterpart. In so doing, we are able to effectively separate the evaluation of the smooth term and the other (potentially non-smooth) term, which are otherwise very expensive to be solved together. As a result, we propose a novel distributed algorithm, termed DuSPA, based on the splitting techniques [25], [26]. Note that the proposed algorithm is able to solve the above problem in a distributed manner without the need of exchanging sensitive information, such as the gradient, and thus is privacy-preserved. Moreover, we characterize the convergence properties for the proposed algorithm under proper assumptions. We now summarize the main contributions as follows:

- 1) Firstly, a novel distributed algorithm is proposed based on splitting methods being applied to the dual counterpart. Different from [6], [8], [20], [21], the proposed algorithm does not require direct communication of the gradient information and is thus suitable for situations where privacy is a key concern. More importantly, another distinctive feature of the proposed algorithm is that it is able to deal with (non-smooth) regularization terms which can represent local constraints using certain indicator function and encode some prior knowledge regarding the solution, e.g., sparsity-prior.
- 2) Secondly, we obtain the convergence properties for the proposed algorithm under proper assumptions. In particular, we establish a specific *non-ergodic* convergence rate of $O(\frac{1}{k})$ for general (potentially non-smooth) convex cost functions, and an explicit linear convergence rate for strongly convex and smooth cost functions. In

contrast to most existing works, we characterize the specific upperbounds of stepsize within which the above convergence rates are attained in terms of the spectral properties of the weight matrix, the Lipschitz constant and strong convexity of the cost function.

Notations: We denote by $x_{i,k}$ the i -th component of a vector $\mathbf{x}$ at time k . In addition, we use ‘1’ to denote all-ones column vector, $\|\cdot\|$ the Euclidean norm of a vector, Δ the difference between two consecutive vectors, i.e., $\Delta \mathbf{x}_{k+1} = \mathbf{x}_{k+1} - \mathbf{x}_k$, $\mathcal{H}$ the Euclidean space, $\langle \cdot, \cdot \rangle$ the inner product, $\bar{\mathcal{R}}$ the extended set of real numbers $\mathcal{R} \cup \{\pm\infty\}$ and $\Gamma_0(\mathcal{H})$ the class of proper lower semi-continuous convex functions from $\mathcal{H}$ to $\bar{\mathcal{R}}$. Moreover, we define a $\mathbf{G}$ -space and its induced norm as $\langle \mathbf{x}, \mathbf{x}' \rangle_{\mathbf{G}} = \langle \mathbf{G}\mathbf{x}, \mathbf{x}' \rangle$ and $\|\mathbf{x}\|_{\mathbf{G}} = \sqrt{\langle \mathbf{G}\mathbf{x}, \mathbf{x} \rangle}$, $\forall \mathbf{x}, \mathbf{x}' \in \mathcal{H}$, where $\mathbf{G}$ is a positive definite matrix. We define the proximal operator¹ as $\text{prox}_{\tau\phi}(\mathbf{v}) = \arg \min_{\mathbf{x} \in \mathcal{H}} \{\phi(\mathbf{x}) + \frac{1}{2\tau} \|\mathbf{x} - \mathbf{v}\|^2\}$ for some convex function $\phi(\cdot)$ and positive constant τ . and the convex conjugate as $\phi^*(\mathbf{y}) := \sup_{\mathbf{x} \in \mathcal{H}} \{\langle \mathbf{x}, \mathbf{y} \rangle - \phi(\mathbf{x})\}$.

II. PROBLEM, ALGORITHM AND ASSUMPTIONS

A. A Motivating Example

Consider a power system consisting of several buses each of which generates certain power within its capacity limit for a specified amount of load. In particular, we have the following economic dispatch problem [28]:

$$\begin{aligned} \min_{\mathbf{q} \in \mathcal{R}^m} \quad & C(\mathbf{q}) = \sum_{i=1}^m C_i(q_i) \\ \text{s.t.} \quad & \sum_i q_i = \sum_i l_i, \\ & \underline{q}_i \leq q_i \leq \bar{q}_i, \quad \forall i \in \mathcal{V} = \{1, 2, \dots, m\}, \end{aligned} \quad (1)$$

where $\mathbf{q} = [q_1, q_2, \dots, q_m]^T$ is the collection vector with q_i being the power produced by the generator associated with bus i , l_i is the load demand from bus i and C_i is the local cost induced by generating certain power. Note that (1) is a canonical economic dispatch problem. Interested readers may refer to [29]–[32] for more detailed models which take into account the power flow dynamics, general (potentially non-convex) cost functions and line flow constraints.

It is not difficult to see that the above economic dispatch problem is a special case of the following general composite resource allocation (optimal exchange) problem (OEP)²:

$$\begin{aligned} \min_{\mathbf{y} \in \mathcal{H}} \quad & \psi(\mathbf{y}) := \sum_{i=1}^m f_i(y_i) + g_i(y_i) \\ \text{s.t.} \quad & \mathbf{1}^T \mathbf{y} = 0, \quad \forall i \in \mathcal{V}, \end{aligned} \quad (2)$$

where $\mathbf{y} = [y_1, y_2, \dots, y_m]^T$ is the collection of all the local decision variables (e.g., power output of generators) to be determined, $f_i : \mathcal{H} \rightarrow \bar{\mathcal{R}}$ is the local cost function which is convex, differentiable and is available only to agent i and $g_i : \mathcal{H} \rightarrow \bar{\mathcal{R}}$ is the regularization term encoding some prior knowledge or certain *feasible* set, which is convex but can be

¹Refer to [27, Def. 12.23] for more details.

²Throughout the paper, without loss of generality, we assume each agent has one-dimension decision variable.

non-smooth. The conventional dual decomposition technique, which requires the information of all decision variables, is not able to solve the above problem in a distributed way. Note that the load l_i can be encoded in the initial value of y_i .

Assumption 1: For each $i \in \mathcal{V}$, the function g_i is proper, closed and convex, i.e., $g_i \in \Gamma_0(\mathcal{H})$, and f_i is convex, differentiable and L_{f_i} -smooth, i.e.,

$$\|\nabla f_i(y'_i) - \nabla f_i(y_i)\| \leq L_{f_i} \|y'_i - y_i\|, \quad \forall y'_i, y_i \in \mathcal{H}. \quad (3)$$

Let $f(\mathbf{y}) := \sum_{i=1}^m f_i(y_i)$, $g(\mathbf{y}) := \sum_{i=1}^m g_i(y_i)$. It is not difficult to see that g is also proper, closed and convex and f has L_f -Lipschitz gradients with $L_f = \max\{L_{f_i}\}$.

B. Communication Network Model

We consider a communication network which can be represented by an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. Each vertex $v_i \in \mathcal{V}$ of the graph denotes an agent while each edge $e_{ij} = (v_i, v_j) \in \mathcal{E}$ a communication link assigned with a positive weight w_{ij} . We use $\mathcal{N}_i := \{j | e_{ij} \in \mathcal{E}\}$ to denote all the neighbors of agent i . In addition, we make the following assumptions on the weight matrix.

Assumption 2: The weight matrix $\mathbf{W}$ associated with the network satisfies the following conditions:

$$(\text{Positive-definiteness}) \quad \mathbf{W}^T = \mathbf{W} \text{ and } \mathbf{W} > 0, \quad (4a)$$

$$(\text{Stochasticity}) \quad \mathbf{W}\mathbf{1} = \mathbf{1} \text{ or } \mathbf{1}^T \mathbf{W} = \mathbf{1}^T, \quad (4b)$$

$$(\text{Connectivity}) \quad \eta := \rho\left(\mathbf{W} - \frac{\mathbf{1}\mathbf{1}^T}{m}\right) < 1. \quad (4c)$$

Remark 1: It follows from the above assumption that $\mathbf{W}$ has a simple eigenvalue one and all the other eigenvalues less than one. It should be noted that $\mathbf{W}$ has to be properly designed such that it is positive-definite, which can be achieved by modified consensus protocols, e.g., $\frac{\mathbf{I} + \mathbf{W}}{2}$ [14].

C. Dual Decomposition

The global Lagrangian for (2) can be written as follows

$$L(\lambda, \mathbf{y}) := \sum_{i=1}^m L_i(\lambda, y_i),$$

where $L_i(\lambda, y_i) := \psi_i(y_i) - \lambda^T y_i$ is the local Lagrangian for agent i with $\psi_i(y_i) := f_i(y_i) + g_i(y_i)$ being the local cost function and λ being the introduced dual variable. Thus, the corresponding dual problem for agent i is

$$D_i(\lambda) := \inf_{y_i} L(\lambda, y_i) = \psi_i^*(\lambda),$$

which, in a global view, is reminiscent of the popular distributed optimization problem proposed in [33]:

$$\min_{\lambda \in \mathcal{H}} \sum_{i=1}^m D_i(\lambda) = \sum_{i=1}^m \psi_i^*(\lambda). \quad (5)$$

It is not difficult to see that the above problem can be rewritten into another decomposable problem by duplicating the dual variable shared by all the agents and introducing

an additional condition that ensures consistency among the introduced auxiliary variables, i.e.,

$$\begin{aligned} \min_{\mathbf{x} \in \mathcal{H}} \quad & \psi^*(\mathbf{x}) = \sum_{i=1}^m \psi_i^*(x_i) \\ \text{s.t.} \quad & x_i = x_j, \forall i, j \in \mathcal{V}, \end{aligned} \quad (6)$$

where $\mathbf{x} = [x_1, x_2, \dots, x_m]^T$ is the collection of all introduced duplicated dual variables.

In [34], it has been shown that, when the topology of the network is connected, the above problem is equivalent to the following optimal consensus problem (OCP):

$$\begin{aligned} \min_{\mathbf{x} \in \mathcal{H}} \quad & \psi^*(\mathbf{x}) = \sum_{i=1}^m (f_i + g_i)^*(x_i) \\ \text{s.t.} \quad & (\mathbf{I} - \mathbf{W})\mathbf{x} = \mathbf{0}. \end{aligned} \quad (7)$$

In the sequel, we term the primal OEP problem and the dual OCP problem together as an OCEP problem for brevity. In addition, for the feasibility of the OCEP problem, we make the following assumption.

Assumption 3: Let $\mathcal{S}$ be the optimal solution set. There exists an optimal solution pair $(\mathbf{x}^*, \mathbf{y}^*) \in \mathcal{S}$ to the OCEP problem.

It was shown in [14] that the optimal solution pair is attained if and only if the following KKT conditions are satisfied:

$$(\text{Primal Feasibility}) \quad \mathbf{1}^T \mathbf{y}^* = 0, \quad (8a)$$

$$(\text{Dual Feasibility}) \quad (\mathbf{I} - \mathbf{W})\mathbf{x}^* = \mathbf{0}, \quad (8b)$$

$$(\text{Lagrangian Optimality}) \quad \mathbf{x}^* - \nabla f(\mathbf{y}^*) \in \partial g(\mathbf{y}^*), \quad (8c)$$

where $\partial g(\mathbf{y}^*)$ denotes the subgradient of g at $\mathbf{y}^*$. As we will show later, with proper initialization, the primal feasibility (8a) is always guaranteed (cf. Lemma 1).

D. Algorithm Development

The essential idea of this paper is, *instead of solving the primal OEP problem directly, to solve the dual OCP problem alternatively in a distributed way*. In particular, to solve the OCP problem, we resort to our previous work [14] which gives the following algorithm:

$$\begin{aligned} \mathbf{x}_{k+1} &= \text{prox}_{\tau \psi^*}(\mathbf{x}_k + \tau \hat{\mathbf{y}}_k), \\ \mathbf{y}_{k+1} &= \mathbf{y}_k - \frac{1}{\tau}(\mathbf{I} - \mathbf{W})\mathbf{x}_{k+1}, \end{aligned} \quad (9)$$

where $\hat{\mathbf{y}}_k = 2\mathbf{y}_k - \mathbf{y}_{k-1}$ and $\tau > 0$ is certain tuning parameter. However, we observe that evaluating the operator $\psi^* = (f + g)^*$ directly is usually expensive and it is not split-table in the conjugate form making it practically impossible for agents to seek the optimum. To make it suitable for distributed processing, we first resort to Moreau identity [26, Th. 14.3]: $\text{prox}_{\tau \psi^*}(\mathbf{s}) = \mathbf{s} - \tau \text{prox}_{\phi/\tau}(\mathbf{s}/\tau)$ for the $\mathbf{x}$ -update of (9), which yields

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{x}_k + \tau \hat{\mathbf{y}}_k - \tau \text{prox}_{\frac{1}{\tau}(f+g)}\left(\frac{\mathbf{x}_k + \tau \hat{\mathbf{y}}_k}{\tau}\right), \\ \mathbf{y}_{k+1} &= \mathbf{y}_k - \frac{1}{\tau}(\mathbf{I} - \mathbf{W})\mathbf{x}_{k+1}. \end{aligned} \quad (10)$$

Algorithm 1 DuSPA for Optimal Exchange Problems

- 1: **Initialization:** $y_{i,0} = 0, \forall i \in \mathcal{V}$ such that $\mathbf{1}^T \mathbf{y}_0 = 0$, and $\mathbf{x}_0, \mathbf{z}_0, \mathbf{y}_{-1}$ arbitrarily assigned.
- 2: **Dual Update:** For $\forall i \in \mathcal{V}$, compute:

$$x_{i,k+1} = x_{i,k} + \tau(\hat{y}_{i,k} - z_{i,k}),$$

where $\hat{y}_{i,k} = 2y_{i,k} - y_{i,k-1}$.

- 3: **Primal Update:** For $\forall i \in \mathcal{V}$, compute

$$y_{i,k+1} = y_{i,k} - \frac{1}{\tau} \sum_{j \in \mathcal{N}_i} w_{ij} (x_{i,k+1} - x_{j,k+1}).$$

- 4: **Update of Auxiliary Variable:** For $\forall i \in \mathcal{V}$, compute

$$z_{i,k+1} = \text{prox}_{\gamma g_i}(z_{i,k} - \gamma \nabla f_i(z_{i,k}) + \gamma \hat{x}_{i+1,k}),$$

where $\hat{x}_{i,k+1} = 2x_{i,k+1} - x_{i,k}$.

- 5: Set $k \rightarrow k+1$ and go to Step 2 until certain stopping criterion is satisfied, e.g., maximum number of iterations.

Let $\mathbf{z}_k = \text{prox}_{\frac{1}{\tau}(f+g)}(\frac{\mathbf{x}_k + \tau \hat{\mathbf{y}}_k}{\tau})$ be the introduced auxiliary variable. Then, we have

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \tau(\hat{\mathbf{y}}_k - \mathbf{z}_k) \in \partial(f+g)(\mathbf{z}_k),$$

which can be written together with the $\mathbf{x}$ -update of (10) as

$$\begin{cases} \mathbf{x}_{k+1} = \mathbf{x}_k + \tau \hat{\mathbf{y}}_k - \tau \mathbf{z}_k \\ \partial(f+g)(\mathbf{z}_k) \ni \mathbf{x}_{k+1} \end{cases} \quad (11)$$

It is noted that $\mathbf{x}_{k+1}$ and $\mathbf{z}_k$ are coupled in the above relation (11), which is still of high computational complexity. We need to decouple the above relation for computational efficiency. To this end, we first relax the second inclusion to $\mathbf{x}_{k+1} \in \partial(f+g)(\mathbf{z}_{k+1})$, which is equivalent to $\min_{\mathbf{z}} f(\mathbf{z}) + g(\mathbf{z}) - \langle \mathbf{z}, \mathbf{x}_{k+1} \rangle$. Then, we use the forward-backward splitting technique [26] to solve its augmented counterpart as follows

$$\min_{\mathbf{z}} f(\mathbf{z}) + g(\mathbf{z}) - \langle \mathbf{z}, \mathbf{x}_{k+1} \rangle + \frac{\epsilon}{2} \|\hat{\mathbf{y}}_k - \mathbf{z}\|^2, \quad (12)$$

where ϵ is a penalty parameter for the proximal term which is to bring some robustness to the above splitting method without impacting the original problem. Thus, we have

$$\mathbf{z}_{k+1} = \text{prox}_{\gamma g}(\mathbf{z}_k - \gamma(\nabla f(\mathbf{z}_k) - \mathbf{x}_{k+1} - \frac{\epsilon}{\tau}(\mathbf{x}_{k+1} - \mathbf{x}_k))),$$

where γ is the stepsize chosen properly for local gradient search. Together with $\mathbf{y}$ -update, and setting $\epsilon = \tau$ we come up with the overall algorithm as follows:

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \tau(\hat{\mathbf{y}}_k - \mathbf{z}_k) \quad (13a)$$

$$\mathbf{y}_{k+1} = \mathbf{y}_k - \frac{1}{\tau}(\mathbf{I} - \mathbf{W})\mathbf{x}_{k+1} \quad (13b)$$

$$\mathbf{z}_{k+1} = \text{prox}_{\gamma g}(\mathbf{z}_k - \gamma(\nabla f(\mathbf{z}_k) - \hat{\mathbf{x}}_{k+1})), \quad (13c)$$

where $\hat{\mathbf{x}}_{k+1} = 2\mathbf{x}_{k+1} - \mathbf{x}_k$. Knowing that $\tau \hat{\mathbf{y}}_k = \tau \mathbf{y}_k + (\mathbf{W} - \mathbf{I})\mathbf{x}_k$, the above equation is equivalent to

$$\mathbf{x}_{k+1} = \mathbf{W}\mathbf{x}_k + \tau(\mathbf{y}_k - \mathbf{z}_k) \quad (14a)$$

$$\tau \mathbf{y}_{k+1} = \tau \mathbf{y}_k - (\mathbf{I} - \mathbf{W})\mathbf{x}_{k+1} \quad (14b)$$

$$\mathbf{z}_{k+1} = \text{prox}_{\gamma g}(\mathbf{z}_k - \gamma(\nabla f(\mathbf{z}_k) - \hat{\mathbf{x}}_{k+1})). \quad (14c)$$

We summarize the proposed algorithm³ in *Algorithm 1*.

Remark 2: Although the above iterate (14c) converges to the solution of (12) only when it runs infinite times, we will show that the algorithm can still converge when it runs only once per iteration. Moreover, the $\mathbf{y}$ -update as given in (14b) resembles the center-free algorithm [8]. Thus, our algorithm can also be regarded as an augmented version of the center-free algorithm that takes into account local constraints. Note that when the regularization term $g(\cdot)$ is an indicator function, the proximal operator $\text{prox}_g(\cdot)$ as shown in (14c) will reduce to a projection operator on certain convex set [26].

Remark 3: It is easy to see that the proposed DuSPA algorithm can be carried out in a distributed manner since each agent only requires local information at each iteration (cf. Step 3 in *Algorithm 1*). In addition, the proposed algorithm is privacy-preserved in that it does not require the direct communication of sensitive gradient information. Further, a proper initialization of the primal variables $\mathbf{y}$ can be simply done in a distributed manner (cf. footnote 3). Of course, some initialization-free techniques can also be employed to avoid the bothering of proper initialization, making it practical in varying environments [21], [22].

Remark 4: It should be noted that although $\mathbf{y}$ and $\mathbf{z}$ are interchangeable in terms of the equilibrium point, we prefer $\mathbf{z}$ as the resource allocation variable when it comes to real physical systems (cf. Section V) since the proximal operator (e.g., physical constraints) is imposed on $\mathbf{z}$ and it is much easier to initialize $\mathbf{y}$ as an auxiliary variable. Further, letting $\mathbf{z}^* = \mathbf{y}^*$, it is straightforward to see that (8a), (8b) and $\mathbf{x}^* - \nabla f(\mathbf{z}^*) \in \partial g(\mathbf{z}^*)$ will constitute another KKT condition set that guarantees the optimality of the solution. It should be also noted that the introduction of $\mathbf{z}$ is essential as it allows us to evaluate the proximal operator while still preserving the conservation property (cf. Lemma 1).

III. SUB-LINEAR CONVERGENCE RATE ANALYSIS

In this section, we study the basic convergence properties of the proposed DuSPA algorithm. In particular, we show that the proposed algorithm has a *non-ergodic* convergence rate of $O(\frac{1}{k})$ for general convex cost functions.

A. Basic Lemmas

We first establish several lemmas which are crucial for the subsequent convergence analysis. Lemma 1 states that the primal variable of DuSPA follows the conservation property. Lemma 2 guarantees the bijective transformation in the disagreement space (i.e., $\text{span}^\perp\{\mathbf{1}\}$). Lemma 3 and 4 provide the basic inequalities for the convergence of DuSPA.

Lemma 1 (Conservation Property [14]): Consider the iterates $\{\mathbf{y}_k\}_{k \geq 0}$ generated by DuSPA. Suppose $\mathbf{1}^T \mathbf{y}_0 = 0$ and Assumption 2 holds, then we have $\mathbf{1}^T \mathbf{y}_k = 0, \forall k \geq 0$.

Lemma 2 (Bijective Transformation [14]): Let $\mathbf{P}$ be a $m \times m$ matrix such that $\text{null}(\mathbf{P}) = \text{span}\{\mathbf{1}\}$. Then, for each $\mathbf{y} \in \text{span}^\perp\{\mathbf{1}\}$, there exists a unique $\mathbf{y}' \in \text{span}^\perp\{\mathbf{1}\}$ such

³Without loss of generality, we assume that the initial demand l_i is zero since otherwise we can simply set $y_{i,0} = l_i$ such that $\mathbf{1}^T \mathbf{y}_0 = \sum_{i=1}^m l_i$.

that $\mathbf{y} = \mathbf{P}\mathbf{y}'$ and vice versa. That is, the $\mathbf{P}$ -transformation between $\mathbf{y} \in \text{span}^\perp\{\mathbf{1}\}$ and $\mathbf{y}' \in \text{span}^\perp\{\mathbf{1}\}$ is bijective.

Lemma 3: Let $\mathbf{p}_k = [\mathbf{x}_k^T, \mathbf{y}_k^T, \mathbf{z}_k^T]^T$ and the sequence $\{\mathbf{p}_k\}_{k \geq 0}$ be generated by the proposed DuSPA algorithm. Suppose Assumptions 1 and 2 hold. If $\gamma < \frac{\lambda_{\min}(\mathbf{W})}{\tau}$, then, for any $k > 0$, we have the following inequality:

$$\|\mathbf{p}_{k+1} - \mathbf{p}^*\|_{\mathbf{H}}^2 - \|\mathbf{p}_k - \mathbf{p}^*\|_{\mathbf{H}}^2 + \|\Delta\mathbf{p}_{k+1}\|_{\mathbf{H}}^2 \leq -2\tau \langle \nabla f(\mathbf{z}_k) - \nabla f(\mathbf{z}^*), \mathbf{z}_{k+1} - \mathbf{z}^* \rangle. \quad (15)$$

Proof: See Appendix. ■

Lemma 4: Suppose Assumptions 1 and 2 hold. If $\gamma < \frac{2\lambda_{\min}(\mathbf{W})}{L_f \lambda_{\min}(\mathbf{W}) + 2\tau}$, then the sequence $\{\|\Delta\mathbf{p}_k\|_{\mathbf{H}}^2\}_{k \geq 0}$ generated by the proposed DuSPA algorithm is monotonically decreasing, i.e.,

$$\|\Delta\mathbf{p}_{k+1}\|_{\mathbf{H}}^2 < \|\Delta\mathbf{p}_k\|_{\mathbf{H}}^2, \quad (16)$$

$$\text{where } \mathbf{H} = \begin{bmatrix} \mathbf{W} & \mathbf{0} & -\tau\mathbf{I} \\ \mathbf{0} & \tau^2\mathbf{Q}^{-1} & \mathbf{0} \\ -\tau\mathbf{I} & \mathbf{0} & \frac{\tau}{\gamma}\mathbf{I} \end{bmatrix}, \mathbf{Q} = \mathbf{I} - (\mathbf{W} - \frac{\mathbf{1}\mathbf{1}^T}{m}).$$

Proof: We first establish a relationship between the sequence of $\|\Delta\mathbf{p}_k\|_{\mathbf{H}}^2$ and the gradient descent of f function at consecutive points. Consider the algorithm (14). Subtracting (14b) from (14a) we obtain $\mathbf{W}\Delta\mathbf{x}_{k+1} = \tau\mathbf{y}_{k+1} - \tau\mathbf{z}_k$. Together with $\mathbf{W}\Delta\mathbf{x}_k = \tau\mathbf{y}_k - \tau\mathbf{z}_{k-1}$, we have

$$\mathbf{W}(\Delta\mathbf{x}_{k+1} - \Delta\mathbf{x}_k) = \tau\Delta\mathbf{y}_{k+1} - \tau\Delta\mathbf{z}_k.$$

Multiplying $\Delta\mathbf{x}_{k+1}$ from both sides we get

$$\begin{aligned} & \langle \mathbf{W}(\Delta\mathbf{x}_{k+1} - \Delta\mathbf{x}_k), \Delta\mathbf{x}_{k+1} \rangle \\ &= \tau \langle \Delta\mathbf{y}_{k+1}, \Delta\mathbf{x}_{k+1} \rangle - \tau \langle \Delta\mathbf{z}_k, \Delta\mathbf{x}_{k+1} \rangle \\ &= \tau \langle \Delta\mathbf{z}_{k+1} - \Delta\mathbf{z}_k, \Delta\mathbf{x}_{k+1} \rangle - \tau \langle \Delta\mathbf{z}_{k+1}, \Delta\mathbf{x}_{k+1} \rangle \\ & \quad + \tau \langle \Delta\mathbf{y}_{k+1}, \Delta\mathbf{x}_{k+1} \rangle. \end{aligned} \quad (17)$$

For the last term of (17), knowing that $\mathbf{1}^T \mathbf{y}_k = 0, \forall k \geq 0$ by Lemma 1, it follows from Lemma 2 that we can find a $\Delta\mathbf{y}'_{k+1} \in \text{span}^\perp\{\mathbf{1}\}$ such that $\Delta\mathbf{y}_{k+1} = (\mathbf{I} - \mathbf{W})\Delta\mathbf{y}'_{k+1}$. Thus, we have

$$\begin{aligned} & \tau \langle \Delta\mathbf{y}_{k+1}, \Delta\mathbf{x}_{k+1} \rangle \\ &= \tau \langle \Delta\mathbf{y}'_{k+1}, (\mathbf{I} - \mathbf{W})\Delta\mathbf{x}_{k+1} \rangle \\ &= \tau^2 \langle \Delta\mathbf{y}_{k+1}, \Delta\mathbf{y}_k - \Delta\mathbf{y}_{k+1} \rangle_{\mathbf{Q}^{-1}}, \end{aligned} \quad (18)$$

where $\mathbf{Q} = \mathbf{I} - (\mathbf{W} - \frac{\mathbf{1}\mathbf{1}^T}{m})$ which is positive-definite by Assumption 2, and we have used the relations $\tau\Delta\mathbf{y}_{k+1} = \tau\Delta\mathbf{y}_k - (\mathbf{I} - \mathbf{W})\Delta\mathbf{x}_{k+1}$ from (14b) and

$$\mathbf{s} = (\mathbf{I} - \mathbf{W})\mathbf{s}' = \mathbf{Q}\mathbf{s}', \forall \mathbf{s}, \mathbf{s}' \in \text{span}^\perp\{\mathbf{1}\} \quad (19)$$

to replace $\Delta\mathbf{y}'_{k+1}$ with $\mathbf{Q}^{-1}\Delta\mathbf{y}_{k+1}$ (cf. Lemma 2).

From (14c), we know that $-\Delta\mathbf{z}_{k+1} + \gamma\hat{\mathbf{x}}_{k+1} - \gamma\nabla f(\mathbf{z}_k) \in \gamma\partial g(\mathbf{z}_{k+1})$ and $-\Delta\mathbf{z}_k + \gamma\hat{\mathbf{x}}_k - \gamma\nabla f(\mathbf{z}_{k-1}) \in \gamma\partial g(\mathbf{z}_k)$. Since ∂g is maximally monotone [26, Def. 20.20] by Assumption 1, letting $\Delta\mathbf{f}_k = \nabla f(\mathbf{z}_k) - \nabla f(\mathbf{z}_{k-1})$, we further have

$$\langle \Delta\mathbf{z}_k - \Delta\mathbf{z}_{k+1} + \gamma\Delta\hat{\mathbf{x}}_{k+1} - \gamma\Delta\mathbf{f}_k, \Delta\mathbf{z}_{k+1} \rangle \geq 0,$$

which is equivalent to

$$\begin{aligned} & \langle \Delta\mathbf{z}_k - \Delta\mathbf{z}_{k+1}, \Delta\mathbf{z}_{k+1} \rangle + \gamma \langle \Delta\mathbf{x}_{k+1} - \Delta\mathbf{x}_k, \Delta\mathbf{z}_{k+1} \rangle \\ & \quad + \gamma \langle \Delta\mathbf{x}_{k+1} - \Delta\mathbf{f}_k, \Delta\mathbf{z}_{k+1} \rangle \geq 0. \end{aligned} \quad (20)$$

Multiplying (20) with $\frac{\tau}{\gamma}$ and combining with (17), (18) yield

$$\begin{aligned} & \langle \Delta\mathbf{x}_{k+1} - \Delta\mathbf{x}_k, \Delta\mathbf{x}_{k+1} \rangle_{\mathbf{W}} + \tau^2 \langle \Delta\mathbf{y}_{k+1} - \Delta\mathbf{y}_k, \Delta\mathbf{y}_{k+1} \rangle_{\mathbf{Q}^{-1}} \\ & \quad - \tau \langle \Delta\mathbf{z}_{k+1} - \Delta\mathbf{z}_k, \Delta\mathbf{x}_{k+1} \rangle + \frac{\tau}{\gamma} \langle \Delta\mathbf{z}_{k+1} - \Delta\mathbf{z}_k, \Delta\mathbf{z}_{k+1} \rangle \\ & \quad - \tau \langle \Delta\mathbf{x}_{k+1} - \Delta\mathbf{x}_k, \Delta\mathbf{z}_{k+1} \rangle + \langle \tau\Delta\mathbf{f}_k, \Delta\mathbf{z}_{k+1} \rangle \leq 0, \end{aligned} \quad (21)$$

which can be rewritten in a compact form as follows

$$\langle \Delta\mathbf{p}_{k+1} - \Delta\mathbf{p}_k, \Delta\mathbf{p}_{k+1} \rangle_{\mathbf{H}} \leq -\tau \langle \Delta\mathbf{f}_k, \Delta\mathbf{z}_{k+1} \rangle. \quad (22)$$

Now, we show the decreasing property of $\|\Delta\mathbf{p}_k\|_{\mathbf{H}}^2$ by using the Lipschitz property of the gradient of f function.

Let $\theta \in (0, 1)$. Then, (22) is equivalent to

$$\begin{aligned} & \langle \Delta\mathbf{p}_{k+1} - \Delta\mathbf{p}_k, \Delta\mathbf{p}_{k+1} \rangle_{\mathbf{H} - \theta \frac{\tau}{\gamma} \mathbf{I}'} \\ & \leq -\tau \langle \Delta\mathbf{f}_k, \Delta\mathbf{z}_{k+1} \rangle - \theta \frac{\tau}{\gamma} \langle \Delta\mathbf{z}_{k+1} - \Delta\mathbf{z}_k, \Delta\mathbf{z}_{k+1} \rangle \\ & = -\theta \frac{\tau}{\gamma} \left\langle \Delta\mathbf{z}_{k+1} - (\Delta\mathbf{z}_k - \frac{\gamma}{\theta} \Delta\mathbf{f}_k), \Delta\mathbf{z}_{k+1} \right\rangle, \end{aligned} \quad (23)$$

where $\mathbf{I}' = \text{diag}(\begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix})$. Then, using $2\langle \mathbf{a}, \mathbf{b} \rangle \leq \|\mathbf{a}\|^2 + \|\mathbf{b}\|^2$ for the last term of (23), we further obtain

$$\begin{aligned} & \left\langle \Delta\mathbf{z}_{k+1} - (\Delta\mathbf{z}_k - \frac{\gamma}{\theta} \Delta\mathbf{f}_k), \Delta\mathbf{z}_{k+1} \right\rangle \\ & \geq \|\Delta\mathbf{z}_{k+1}\|^2 - \frac{1}{2} \|\Delta\mathbf{z}_{k+1}\|^2 - \frac{1}{2} \left\| \Delta\mathbf{z}_k - \frac{\gamma}{\theta} \Delta\mathbf{f}_k \right\|^2 \\ & = \frac{1}{2} \|\Delta\mathbf{z}_{k+1}\|^2 - \frac{1}{2} \left\| \Delta\mathbf{z}_k - \frac{\gamma}{\theta} \Delta\mathbf{f}_k \right\|^2. \end{aligned} \quad (24)$$

Moreover, knowing that $\frac{1}{L_f} \|\Delta\mathbf{f}_k\|^2 \leq \langle \Delta\mathbf{z}_k, \Delta\mathbf{f}_k \rangle$ by the convexity of f and the Lipschitz assumption (cf. Assumption 1 and [35, p.24]), we have

$$\begin{aligned} & \left\| \Delta\mathbf{z}_k - \frac{\gamma}{\theta} \Delta\mathbf{f}_k \right\|^2 \\ &= \|\Delta\mathbf{z}_k\|^2 + \frac{\gamma^2}{\theta^2} \|\Delta\mathbf{f}_k\|^2 - 2\frac{\gamma}{\theta} \langle \Delta\mathbf{z}_k, \Delta\mathbf{f}_k \rangle \\ & \leq \|\Delta\mathbf{z}_k\|^2 + \frac{\gamma}{\theta^2} (\gamma - \frac{2\theta}{L_f}) \|\Delta\mathbf{f}_k\|^2. \end{aligned} \quad (25)$$

Let $\gamma < \frac{2\theta}{L_f}$. Combining (24) and (25), we obtain

$$\begin{aligned} & \left\langle \Delta\mathbf{z}_{k+1} - (\Delta\mathbf{z}_k - \frac{\gamma}{\theta} \Delta\mathbf{f}_k), \Delta\mathbf{z}_{k+1} \right\rangle \\ & \geq \frac{1}{2} \|\Delta\mathbf{z}_{k+1}\|^2 - \frac{1}{2} (\|\Delta\mathbf{z}_k\|^2 + \frac{\gamma}{\theta^2} (\gamma - \frac{2\theta}{L_f}) \|\Delta\mathbf{f}_k\|^2) \\ & > \frac{1}{2} \|\Delta\mathbf{z}_{k+1}\|^2 - \frac{1}{2} \|\Delta\mathbf{z}_k\|^2. \end{aligned} \quad (26)$$

Let $\gamma < \frac{(1-\theta)\lambda_{\min}(\mathbf{W})}{\tau}$. Since $\mathbf{Q} > 0, \mathbf{W} > 0$, we further have

$$(1-\theta) \frac{\tau}{\gamma} \mathbf{I} - [\tau\mathbf{I} \quad \mathbf{0}] \begin{bmatrix} \mathbf{W}^{-1} & \mathbf{0} \\ \mathbf{0} & \frac{1}{\tau^2} \mathbf{Q} \end{bmatrix} \begin{bmatrix} \tau\mathbf{I} \\ \mathbf{0} \end{bmatrix} > 0. \quad (27)$$

It follows from Schur complement that $\mathbf{H} - \theta \frac{\tau}{\gamma} \mathbf{I}' > 0$. Thus, using (26) and $2\langle \mathbf{a}, \mathbf{b} \rangle_{\mathbf{G}} \leq \|\mathbf{a}\|_{\mathbf{G}}^2 + \|\mathbf{b}\|_{\mathbf{G}}^2$ in $\mathbf{G}$ -space, the relation (23) can be rewritten as

$$\begin{aligned} & \|\Delta\mathbf{p}_{k+1}\|_{\mathbf{H} - \theta \frac{\tau}{\gamma} \mathbf{I}'}^2 - \|\Delta\mathbf{p}_k\|_{\mathbf{H} - \theta \frac{\tau}{\gamma} \mathbf{I}'}^2 \\ & \leq 2 \langle \Delta\mathbf{p}_{k+1} - \Delta\mathbf{p}_k, \Delta\mathbf{p}_{k+1} \rangle_{\mathbf{H} - \theta \frac{\tau}{\gamma} \mathbf{I}'} \\ & < -\theta \frac{\tau}{\gamma} (\|\Delta\mathbf{z}_{k+1}\|^2 - \|\Delta\mathbf{z}_k\|^2) \\ & = -\|\Delta\mathbf{p}_{k+1}\|_{\theta \frac{\tau}{\gamma} \mathbf{I}'}^2 + \|\Delta\mathbf{p}_k\|_{\theta \frac{\tau}{\gamma} \mathbf{I}'}^2. \end{aligned} \quad (28)$$

Setting $\theta = \frac{L_f \lambda_{\min}(\mathbf{W})}{2\tau + L_f \lambda_{\min}(\mathbf{W})}$ such that $\gamma < \frac{2\lambda_{\min}(\mathbf{W})}{L_f \lambda_{\min}(\mathbf{W}) + 2\tau} = \min\{\frac{2\theta}{L_f}, \frac{(1-\theta)\lambda_{\min}(\mathbf{W})}{\tau}\}$, (26) and (28) are both valid. Then, rearranging the terms of (28) completes the proof. ■

B. Main Result

Theorem 5: Suppose Assumptions 1, 2 and 3 hold. If $\gamma < \frac{\lambda_{\min}(\mathbf{W})}{L_f \lambda_{\min}(\mathbf{W}) + \tau}$, then the sequence $\{(\mathbf{x}_k, \mathbf{y}_k)\}_{k \geq 0}$ generated by the proposed DuSPA algorithm will converge to an optimal solution pair $(\mathbf{x}^*, \mathbf{y}^*)$ of the OCEP problem. Moreover, the fixed point residual in terms of $\|\Delta \mathbf{p}_{k+1}\|_{\mathbf{H}}^2$ will decrease at a non-ergodic rate of $O(\frac{1}{k})$. In particular, we have

$$\|\Delta \mathbf{p}_{k+1}\|_{\mathbf{H}}^2 < \frac{(\lambda_{\min}(\mathbf{W}) - \gamma\tau) \|\mathbf{p}_0 - \mathbf{p}^*\|_{\mathbf{H}}^2}{k L_f \gamma \lambda_{\min}(\mathbf{W})}.$$

Proof: For any $(\mathbf{x}^*, \mathbf{y}^*) \in \mathcal{S}$ and $\mathbf{z}^* = \mathbf{y}^*$, using Lemma 3 and the convexity of f , we have

$$\begin{aligned} \|\mathbf{p}_{k+1} - \mathbf{p}^*\|_{\mathbf{H}}^2 - \|\mathbf{p}_k - \mathbf{p}^*\|_{\mathbf{H}}^2 + \|\Delta \mathbf{p}_{k+1}\|_{\mathbf{H}}^2 \\ \leq -2\tau \langle \nabla f(\mathbf{z}_k) - \nabla f(\mathbf{z}^*), \mathbf{z}_{k+1} - \mathbf{z}^* \rangle \\ + 2\tau \langle \nabla f(\mathbf{z}_{k+1}) - \nabla f(\mathbf{z}^*), \mathbf{z}_{k+1} - \mathbf{z}^* \rangle \\ \leq 2\tau \langle \nabla f(\mathbf{z}_{k+1}) - \nabla f(\mathbf{z}_k), \mathbf{z}_{k+1} - \mathbf{z}^* \rangle \\ = 2\tau [D_f(\mathbf{z}^*, \mathbf{z}_{k+1}) - D_f(\mathbf{z}^*, \mathbf{z}_k) + D_f(\mathbf{z}_{k+1}, \mathbf{z}_k)], \end{aligned} \quad (29)$$

where $D_f(\mathbf{x}, \mathbf{x}') = f(\mathbf{x}) - f(\mathbf{x}') - \langle \nabla f(\mathbf{x}'), \mathbf{x} - \mathbf{x}' \rangle$, $\forall \mathbf{x}, \mathbf{x}' \in \mathcal{H}$ is the Bregman distance of f evaluated at $\mathbf{x}'$.

Let $V_k = \|\mathbf{p}_k - \mathbf{p}^*\|_{\mathbf{H}}^2 - 2\tau D_f(\mathbf{z}^*, \mathbf{z}_k)$. Rearranging terms of (29), we further have

$$V_k - V_{k+1} \geq \|\Delta \mathbf{p}_{k+1}\|_{\mathbf{H}}^2 - 2\tau D_f(\mathbf{z}_{k+1}, \mathbf{z}_k). \quad (30)$$

Since ∇f is L_f -Lipschitz continuous, we have

$$2\tau D_f(\mathbf{z}_{k+1}, \mathbf{z}_k) \leq \tau L_f \|\Delta \mathbf{z}_{k+1}\|^2 = \tau L_f \|\Delta \mathbf{p}_{k+1}\|_{\mathbf{I}'}^2.$$

Now, we show that there exists a proper constant such that the value of V_k in (30) is decreasing over time.

Let $c = \frac{\gamma \lambda_{\min}(\mathbf{W})}{\lambda_{\min}(\mathbf{W}) - \gamma\tau}$. Then, we have

$$\left(\frac{\tau}{\gamma} - \frac{\tau}{c}\right) \mathbf{I} - \begin{bmatrix} \tau \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \frac{1}{\tau^2} \mathbf{Q}^{-1} \end{bmatrix} \begin{bmatrix} \tau \mathbf{I} \\ \mathbf{0} \end{bmatrix} > \mathbf{0}.$$

By Schur complement lemma, we have $\mathbf{H} - \frac{\tau}{c} \mathbf{I}' > \mathbf{0}$ and thus $\tau L_f \mathbf{I}' < c L_f \mathbf{H}$. Then, (30) can be rewritten as

$$V_k - V_{k+1} \geq \|\Delta \mathbf{p}_{k+1}\|_{\mathbf{H} - \tau L_f \mathbf{I}'}^2 > (1 - c L_f) \|\Delta \mathbf{p}_{k+1}\|_{\mathbf{H}}^2. \quad (31)$$

Since $\gamma < \frac{\lambda_{\min}(\mathbf{W})}{L_f \lambda_{\min}(\mathbf{W}) + \tau}$, we have $c L_f < 1$ and $\mathbf{H} - \tau L_f \mathbf{I}' > \mathbf{0}$. Thus, we can obtain that $V_k \geq \|\mathbf{p}_k - \mathbf{p}^*\|_{\mathbf{H} - \tau L_f \mathbf{I}'}^2 > 0, \forall k$. Then, summing (31) from 1 to t over k we have

$$\sum_{k=1}^t \|\Delta \mathbf{p}_{k+1}\|_{\mathbf{H}}^2 < \frac{V_0}{1 - c L_f}. \quad (32)$$

Also, knowing that $\gamma < \frac{\lambda_{\min}(\mathbf{W})}{L_f \lambda_{\min}(\mathbf{W}) + \tau} < \frac{2\lambda_{\min}(\mathbf{W})}{L_f \lambda_{\min}(\mathbf{W}) + 2\tau}$, it follows from Lemma 4 that $\|\Delta \mathbf{p}_{t+1}\|_{\mathbf{H}}^2$ is monotonically decreasing. Thus, we further have

$$\|\Delta \mathbf{p}_{t+1}\|_{\mathbf{H}}^2 < \frac{1}{t} \sum_{k=1}^t \|\Delta \mathbf{p}_{k+1}\|_{\mathbf{H}}^2 < \frac{V_0}{t(1 - c L_f)} \quad (33)$$

and that $\lim_{t \rightarrow \infty} \|\Delta \mathbf{p}_{t+1}\|_{\mathbf{H}} = 0$.

⁴From (31), we know that V_k is bounded and so is $\mathbf{x}_k$ and $\mathbf{y}_k$. Thus, the sequence $\{(\mathbf{x}_k, \mathbf{y}_k)\}_{k \geq 0}$ has at least a cluster point $(\mathbf{x}_{\infty}, \mathbf{y}_{\infty})$ and the corresponding subsequence $\{(x_{k_i}, y_{k_i})\}_{i \geq 0}$. Taking the limit along $k_i \rightarrow \infty$ in (14), we know that the cluster point $(\mathbf{x}_{\infty}, \mathbf{y}_{\infty})$ is an optimal solution pair of the OCEP problem. Since $\{V_k\}_{k \geq 0}$ converges and is contractive with respect to the optimal set $\mathcal{S}$, (31) implies that $(\mathbf{x}_{\infty}, \mathbf{y}_{\infty})$ is the unique cluster point (cf. [26, Th. 5.5] and [36, Th. 3.7]). It then follows that $\{(\mathbf{x}_k, \mathbf{y}_k)\}_{k \geq 0}$ will converge to a unique point of the optimal set $\mathcal{S}$. Thus, we conclude that the sequence $\{(\mathbf{x}_k, \mathbf{y}_k)\}_{k \geq 0}$ will converge to a solution pair $(\mathbf{x}^*, \mathbf{y}^*)$ with a non-ergodic convergence rate of $O(\frac{1}{k})$. ■

Remark 5: When τ is chosen sufficiently small, there will be two time-scale separation between the update of x and z . As a result, the algorithm will resemble the centralized counterpart and the upperbound of γ will also reach the similar bound as implied by centralized counterparts (cf. Theorem 4). However, small value of τ leads to a slow convergence rate. It should be noted that we provide a very simple upperbound of γ which is not available in the existing literature.

IV. LINEAR CONVERGENCE RATE ANALYSIS

In this section, we show that if the cost function f is smooth and strongly convex and $g = 0$, the proposed DuSPA algorithm will converge linearly to the optimum.

Assumption 4: For each $i \in \mathcal{V}$, the cost function f_i is l_{f_i} -strongly convex, i.e.,

$$\|\nabla f_i(y'_i) - \nabla f_i(y_i)\| \geq l_{f_i} \|y'_i - y_i\|, \quad \forall y'_i, y_i \in \mathcal{H}. \quad (34)$$

It is not difficult to see that the global cost function f is l_f -strongly convex with $l_f = \min\{l_{f_i}\}$.

Theorem 6: Let $\varepsilon = \gamma\tau$. Suppose Assumptions 1, 2, 3 and 4 hold. If $\gamma < \frac{\lambda_{\min}(\mathbf{W})}{4(L_f + l_f)\lambda_{\min}(\mathbf{W}) + 5\tau}$ and $\tau < \frac{L_f l_f}{L_f + l_f}$, then the sequence $\{(\mathbf{x}_k, \mathbf{y}_k)\}_{k \geq 0}$ generated by the proposed DuSPA algorithm will converge linearly to an optimal solution pair $(\mathbf{x}^*, \mathbf{y}^*)$ of the OCEP problem. In particular, we have

$$\|\mathbf{p}_{k+1} - \mathbf{p}^*\|_{\mathbf{H}'}^2 \leq (1 - \delta) \|\mathbf{p}_k - \mathbf{p}^*\|_{\mathbf{H}'}^2, \quad (35)$$

where $\mathbf{H}' = \begin{bmatrix} \varepsilon \mathbf{I} + \mathbf{W} & \mathbf{0} & -\tau \mathbf{I} \\ \mathbf{0} & \tau^2(\varepsilon \mathbf{I} + \mathbf{Q}^{-1}) & \mathbf{0} \\ -\tau \mathbf{I} & \mathbf{0} & \frac{\tau}{\gamma} \mathbf{I} \end{bmatrix}$ and

$$\delta = \min \left\{ \frac{\varepsilon}{4\varepsilon + 2}, \gamma \left(\frac{L_f l_f}{L_f + l_f} - \tau \right), \frac{\varepsilon(1 - \eta)}{2 - \eta} \right\}.$$

Proof: We first establish the basic inequality showing the decreasing property of the optimality gap $\|\mathbf{p}_k - \mathbf{p}^*\|_{\mathbf{H}}$ based on the smoothness and strong convexity of f . For any $(\mathbf{x}^*, \mathbf{y}^*) \in \mathcal{S}$ and $\mathbf{z}^* = \mathbf{y}^*$, using Lemma 3 we obtain

$$\begin{aligned} \|\mathbf{p}_{k+1} - \mathbf{p}^*\|_{\mathbf{H}}^2 - \|\mathbf{p}_k - \mathbf{p}^*\|_{\mathbf{H}}^2 \\ \leq -\|\Delta \mathbf{p}_{k+1}\|_{\mathbf{H}}^2 - 2\tau \langle \nabla f(\mathbf{z}_k) - \nabla f(\mathbf{z}^*), \mathbf{z}_{k+1} - \mathbf{z}^* \rangle \\ \leq -\|\Delta \mathbf{p}_{k+1}\|_{\mathbf{H}}^2 - 2\tau \langle \nabla f(\mathbf{z}_k) - \nabla f(\mathbf{z}^*), \mathbf{z}_k - \mathbf{z}^* \rangle \\ - 2\tau \langle \nabla f(\mathbf{z}_k) - \nabla f(\mathbf{z}^*), \Delta \mathbf{z}_{k+1} \rangle. \end{aligned} \quad (36)$$

⁴Similar argument has appeared in our previous work [15]. We include it here for completeness of the proof.

For the last term, using Young's inequality, for any given positive number β we have

$$\begin{aligned} & \langle \nabla f(\mathbf{z}_k) - \nabla f(\mathbf{z}^*), \Delta \mathbf{z}_{k+1} \rangle \\ & \geq -\frac{\|\nabla f(\mathbf{z}_k) - \nabla f(\mathbf{z}^*)\|^2}{2\beta} - \frac{\beta \|\Delta \mathbf{z}_{k+1}\|^2}{2}. \end{aligned} \quad (37)$$

Since f is L_f -smooth and l_f -strongly convex, for the second term of (36) we have

$$\begin{aligned} & \langle \nabla f(\mathbf{z}_k) - \nabla f(\mathbf{z}^*), \mathbf{z}_k - \mathbf{z}^* \rangle \\ & \geq \frac{1}{L_f + l_f} \|\nabla f(\mathbf{z}_k) - \nabla f(\mathbf{z}^*)\|^2 + \frac{l_f L_f}{L_f + l_f} \|\mathbf{z}_k - \mathbf{z}^*\|^2. \end{aligned} \quad (38)$$

Combining (36), (37) and (38) leads to

$$\begin{aligned} & \|\mathbf{p}_{k+1} - \mathbf{p}^*\|_{\mathbf{H}}^2 - \|\mathbf{p}_k - \mathbf{p}^*\|_{\mathbf{H}}^2 \\ & \leq -\|\Delta \mathbf{p}_{k+1}\|_{\mathbf{H}}^2 + \tau\beta \|\Delta \mathbf{z}_{k+1}\|^2 - \frac{2l_f L_f \tau}{L_f + l_f} \|\mathbf{z}_k - \mathbf{z}^*\|^2 \\ & \quad - \left(\frac{2\tau}{L_f + l_f} - \frac{\tau}{\beta} \right) \|\nabla f(\mathbf{z}_k) - \nabla f(\mathbf{z}^*)\|^2. \end{aligned} \quad (39)$$

Then, we show that the magnitude of the decrease of the optimality gap is at least proportional to itself. Recalling that $\mathbf{y}^* = \mathbf{z}^*$, using (54) we have

$$W\Delta \mathbf{x}_{k+1} + \tau(\mathbf{z}_k - \mathbf{z}^*) = \tau(\mathbf{y}_{k+1} - \mathbf{y}^*)$$

which, taking the norms of both sides of the above equation and applying the basic inequality $\|\mathbf{a} + \mathbf{b}\|^2 \leq (1 + \frac{1}{\mu})\|\mathbf{a}\|^2 + (1 + \mu)\|\mathbf{b}\|^2$, $\forall \mu > 0$, yields that

$$\begin{aligned} & \tau^2 \|\mathbf{y}_{k+1} - \mathbf{y}^*\|^2 \\ & \leq (1 + \mu) \|\Delta \mathbf{x}_{k+1}\|_{\mathbf{W}^2}^2 + \tau^2 \left(1 + \frac{1}{\mu}\right) \|\mathbf{z}_k - \mathbf{z}^*\|^2. \end{aligned} \quad (40)$$

Adding $-\tau^2 \|\mathbf{y}_k - \mathbf{y}^*\|^2$ to both sides of (40), we further have

$$\begin{aligned} & \tau^2 \|\mathbf{y}_{k+1} - \mathbf{y}^*\|^2 - \tau^2 \|\mathbf{y}_k - \mathbf{y}^*\|^2 \\ & \leq (1 + \mu) \|\Delta \mathbf{x}_{k+1}\|_{\mathbf{W}^2}^2 + \tau^2 \left(1 + \frac{1}{\mu}\right) \|\mathbf{z}_k - \mathbf{z}^*\|^2 \\ & \quad - \tau^2 \|\mathbf{y}_k - \mathbf{y}^*\|^2. \end{aligned} \quad (41)$$

Likewise, knowing that $g = 0$, using (14c) we have

$$-\Delta \mathbf{z}_{k+1} + \gamma \hat{\mathbf{x}}_{k+1} - \gamma \nabla f(\mathbf{z}_k) = 0.$$

Together with the optimality condition (8c) and the fact that $\mathbf{z}^* = \mathbf{y}^*$, the above equality can be rewritten as follows

$$\gamma(\hat{\mathbf{x}}_{k+1} - \mathbf{x}^*) = \gamma(\nabla f(\mathbf{z}_k) - \nabla f(\mathbf{z}^*)) + \Delta \mathbf{z}_{k+1}.$$

Thus, using the same way to obtain (40), we have

$$\begin{aligned} \|\hat{\mathbf{x}}_{k+1} - \mathbf{x}^*\|^2 & = \left\| \nabla f(\mathbf{z}_k) - \nabla f(\mathbf{z}^*) + \frac{1}{\gamma}(\Delta \mathbf{z}_{k+1}) \right\|^2 \\ & \leq (1 + \nu) \|\nabla f(\mathbf{z}_k) - \nabla f(\mathbf{z}^*)\|^2 \\ & \quad + \left(1 + \frac{1}{\nu}\right) \frac{1}{\gamma^2} \|\Delta \mathbf{z}_{k+1}\|^2, \end{aligned} \quad (42)$$

where ν is some positive number. Further, employing the parallelogram identity $\|\mathbf{a} + \mathbf{b}\|^2 + \|\mathbf{a} - \mathbf{b}\|^2 = 2\|\mathbf{a}\|^2 + 2\|\mathbf{b}\|^2$ for the left-hand side of (42), we have

$$\begin{aligned} \|\hat{\mathbf{x}}_{k+1} - \mathbf{x}^*\|^2 & = 2\|\mathbf{x}_{k+1} - \mathbf{x}^*\|^2 + 2\|\Delta \mathbf{x}_{k+1}\|^2 - \|\mathbf{z}_k - \mathbf{z}^*\|^2. \end{aligned} \quad (43)$$

Combining (42) and (43) and subtracting $\|\mathbf{x}_k - \mathbf{x}^*\|^2$ from both sides of the obtained inequality yield

$$\begin{aligned} & 2\|\mathbf{x}_{k+1} - \mathbf{x}^*\|^2 - 2\|\mathbf{x}_k - \mathbf{x}^*\|^2 \\ & \leq (1 + \nu) \|\nabla f(\mathbf{z}_k) - \nabla f(\mathbf{z}^*)\|^2 + \left(1 + \frac{1}{\nu}\right) \frac{1}{\gamma^2} \|\Delta \mathbf{z}_{k+1}\|^2 \\ & \quad - 2\|\Delta \mathbf{x}_{k+1}\|^2 - \|\mathbf{x}_k - \mathbf{x}^*\|^2. \end{aligned} \quad (44)$$

Multiplying both sides of (41) by ε and (44) by $\frac{\varepsilon}{2}$ and then combining it with (39), we obtain (in a compact form)

$$\begin{aligned} & \|\mathbf{p}_{k+1} - \mathbf{p}^*\|_{\mathbf{H}'}^2 - \|\mathbf{p}_k - \mathbf{p}^*\|_{\mathbf{H}'}^2 \\ & \leq -\|\Delta \mathbf{p}_{k+1}\|_{\mathbf{H} - \mathbf{A}_1}^2 - \|\mathbf{p}_k - \mathbf{p}^*\|_{\mathbf{A}_2}^2 \\ & \quad - \left(\frac{2\tau}{L_f + l_f} - \frac{\tau}{\beta} - \frac{(1 + \nu)\varepsilon}{2} \right) \|\nabla f(\mathbf{z}_k) - \nabla f(\mathbf{z}^*)\|^2, \end{aligned} \quad (45)$$

where

$$\mathbf{A}_1 = \begin{bmatrix} \varepsilon((1 + \mu)\mathbf{W}^2 - \mathbf{I}) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \left(\frac{1}{2}\left(1 + \frac{1}{\nu}\right)\frac{\tau}{\gamma} + \tau\beta\right)\mathbf{I} \end{bmatrix}$$

and

$$\mathbf{A}_2 = \begin{bmatrix} \frac{\varepsilon}{2}\mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \varepsilon\tau^2\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \left(\frac{2L_f l_f \tau}{L_f + l_f} - \left(1 + \frac{1}{\mu}\right)\varepsilon\tau^2\right)\mathbf{I} \end{bmatrix}.$$

Now, we establish the specific conditions on which the linear convergence rate is attained. To this end, it suffices to show that there exists a positive number $0 < \delta < 1$ such that (35) is valid. In view of (45), this amounts to showing that

$$\mathbf{H} > \mathbf{A}_1, \quad (46)$$

$$\frac{2\tau}{L_f + l_f} - \frac{\tau}{\beta} - \frac{1}{2}(1 + \nu)\varepsilon \geq 0, \quad (47)$$

$$\mathbf{A}_2 > \delta \mathbf{H}'. \quad (48)$$

In the following, we will show that the above three inequalities are satisfied for $\mu = 1$, $\nu = 2$.

Let us first consider (46). Taking

$$\gamma < \min \left\{ \frac{\lambda_2}{\tau}, \frac{\lambda_2}{4\lambda_2\beta + 5\tau} \right\},$$

where we denote $\lambda_2 = \lambda_{\min}(\mathbf{W})$, and recalling that $\varepsilon = \gamma\tau$, one immediately has

$$\varepsilon\mathbf{I} + \mathbf{W} - \varepsilon(1 + \mu)\mathbf{W}^2 > 0. \quad (49)$$

Moreover, it is straightforward to show that

$$\gamma < \frac{5\tau + 4\beta\lambda_2 - \sqrt{(5\tau + 4\beta\lambda_2)^2 - 16\beta\tau\lambda_2}}{8\beta\tau}.$$

This means that

$$4\beta\tau\gamma^2 - (5\tau + 4\beta\lambda_2)\gamma + \lambda_2 > 0,$$

which is equivalent to the following relation

$$\frac{1}{4\gamma} - \beta > \frac{\tau}{\lambda_2 - \varepsilon}.$$

Noting that $\lambda_{\max}(\mathbf{W}) = 1$, we have $\lambda_{\min}(\varepsilon\mathbf{I} + \mathbf{W} - \varepsilon(1 + \mu)\mathbf{W}^2) = \lambda_{\min}(\mathbf{W}) - \varepsilon\mu$. Since $\mu = 1, \nu = 2$, it follows from the above relation that

$$\left(\frac{\tau}{\gamma} - \frac{1}{2} \left(1 + \frac{1}{\nu}\right) \frac{\tau}{\gamma} - \tau\beta\right) \mathbf{I} - \tau^2(\varepsilon\mathbf{I} + \mathbf{W} - \varepsilon(1 + \mu)\mathbf{W}^2)^{-1} > 0. \quad (50)$$

Combining (49) and (50), and using Schur complement lemma, one can obtain $\mathbf{H} > \mathbf{\Lambda}_1$, which is (46).

Next, let us consider (47). We can take

$$\gamma < \frac{2}{3} \left(\frac{2}{L_f + l_f} - \frac{1}{\beta} \right)$$

to obtain (47) for $\mu = 1$ and $\nu = 2$.

To sum up, we can take $\beta = L_f + l_f$, $\mu = 1$ and $\nu = 2$ for (37), (41) and (44), respectively, so that

$$\gamma < \min \left\{ \frac{\lambda_2}{\tau}, \frac{\lambda_2}{4\lambda_2(L_f + l_f) + 5\tau}, \frac{2}{3} \frac{1}{L_f + l_f} \right\}$$

and the relations (46) and (47) are satisfied.

It remains to consider the last relation (48). We claim that

$$\delta \leq \min \left\{ \frac{\varepsilon}{4\varepsilon + 2}, \gamma \left(\frac{L_f l_f}{L_f + l_f} - \tau \right), \frac{\varepsilon(1 - \eta)}{2 - \eta} \right\}$$

with $\tau < \frac{L_f l_f}{L_f + l_f}$ satisfies (48). To see this, we first note that if $\delta < \frac{\varepsilon}{4\varepsilon + 2}$, then $\frac{2\delta\varepsilon}{\varepsilon(1 - 2\delta) - 2\delta} < 1$. It further implies

$$\frac{2L_f l_f \tau}{L_f + l_f} - \frac{\tau\delta}{\gamma} - \left(1 + \frac{1}{\mu}\right) \tau^2 \geq \frac{\delta\tau}{\gamma} > \frac{2\delta^2\tau^2}{\varepsilon(1 - 2\delta) - 2\delta},$$

which, together with the fact that $\lambda_{\max}(\mathbf{W}) = 1$, yields

$$\left(\frac{2L_f l_f \tau}{L_f + l_f} - \frac{\tau\delta}{\gamma} - \left(1 + \frac{1}{\mu}\right) \tau^2 \right) \mathbf{I} - \delta^2\tau^2 \left(\frac{1}{2}\varepsilon\mathbf{I} - \delta(\varepsilon\mathbf{I} + \mathbf{W}) \right)^{-1} > 0. \quad (51)$$

Moreover, it is not difficult to verify that for $\delta \leq \min \left\{ \frac{\varepsilon}{4\varepsilon + 2}, \frac{\varepsilon(1 - \eta)}{2 - \eta} \right\}$, we have

$$\varepsilon\mathbf{I} - 2\delta(\varepsilon\mathbf{I} + \mathbf{W}) > 0, \quad (52)$$

and

$$(\varepsilon - \delta)\tau^2\mathbf{I} - \delta\tau^2\mathbf{Q}^{-1} > 0. \quad (53)$$

Therefore, applying Schur complement lemma to (51), (52) and (53) completes the proof of the aforementioned claim, from which the theorem follows. ■

V. APPLICATIONS TO DISTRIBUTED ECONOMIC DISPATCH

We report a simulation result to show the effectiveness of the proposed algorithm. We consider an economic dispatch problem as described in Section II-A. The cost induced by producing certain amount of power can be approximately modeled as a quadratic function [28]:

$$C(q) = aq^2 + bq + c,$$

where a, b, c are some coefficients and q is the power output of generators. Each generator is also subject to certain capacity

limit, i.e., $q \in [\underline{q}, \bar{q}]$. We assume each bus is associated with a generator and certain load known locally by itself. If a bus has no generator, without loss of generality, we can simply set $a = b = c = 0$ and $\bar{q} = \underline{q} = 0$. Note that this setting could also be used to model the plug-and-play of a generator.

In relation to our algorithm, the variable $\mathbf{x}$ can be regarded as the incremental cost while $\mathbf{z}$ denotes the power output of each generator and $\mathbf{y}$ represents the *virtual* power demand required locally from all buses. When an optimal point is attained, the incremental cost $\mathbf{x}$ will reach a consensus value and $\mathbf{z}$ gives the optimal output of generators.

A. Simulation Setting

We consider an IEEE 14-bus power system. The parameters of the generators are adopted from [18] and restated in Table I. We assume the communication network is the same with the coupling structure of buses. The weight matrix is designed based on constant-weight⁵ protocol [37], i.e., $\mathbf{I} - \alpha\mathbf{L}$, where $\mathbf{L}$ is the Laplacian matrix of the graph and α is some small number. The algorithm is properly initialized to the optimal point corresponding to a total demand of 231MW which is distributed among buses. We then increase⁶ the level of demand by 60% for each bus to mimic the change of load which results in some generator reaching its capacity limit.

TABLE I
GENERATOR PARAMETERS

Bus	a (\$/MW ²)	b (\$/MW)	c (\$)	$[p, \bar{p}]$ (MW)
1	0.04	2.0	0	[0, 80]
2	0.03	3.0	0	[0, 90]
3	0.035	4.0	0	[0, 70]
6	0.03	4.0	0	[0, 70]
8	0.04	2.5	0	[0, 80]

B. Comparing with the State-of-the-Art

We compare the proposed DuSPA algorithm with the state-of-the-art algorithm, termed Distributed Incremental Cost Consensus (D-ICC) [19], in terms of the overall power mismatch under different parameter settings. Note that the new optimal point is attained when the power output of the generators are $z_1^* = 78.43\text{MW}$, $z_2^* = 87.91\text{MW}$, $z_3^* = 61.06\text{MW}$, $z_6^* = 70\text{MW}$, $z_8^* = 72.18\text{MW}$, respectively.

Fig. 1 plots the total power mismatch of the IEEE 14-bus system when subjecting to certain demand change under the proposed DuSPA algorithm and the state-of-the-art D-ICC algorithm respectively. It follows from the figure that D-ICC has similar performance with DuSPA when $\tau/\varepsilon = 0.005^7$, $\gamma = 5$ while DuSPA outperforms D-ICC with bigger values of τ . For instance, DuSPA converges faster than D-ICC when $\tau/\varepsilon = 0.05$ while D-ICC diverges at $\varepsilon = 0.06$. It should be noted that the convergence of D-ICC for large stepsizes is not theoretically guaranteed according to [19]. Fig. 2 plots the

⁵The modified Metropolis-Hastings protocol can be also employed [14].

⁶This is done by adding $0.6 \cdot l$ to the initial value of y_1 .

⁷This parameter setting is suggested in [19, Section V-H].

Fig. 1. Plot of the historical evolution of the total power mismatch (in a log scale) of the IEEE 14-bus system in response to the demand change under the proposed DuSPA algorithm and the state-of-the-art D-ICC algorithm with different parameter settings, respectively.

Fig. 2. The historical evolution of the power output of the generators at Buses 1, 2, 3, 6 and 8 under the proposed DuSPA algorithm with $\tau = 0.01, \gamma = 5$. The generator at Bus 6 has reached to its capacity limit, i.e., 70MW.

historical evolution of the power output of the generators at Buses 1, 2, 3, 6 and 8 under the proposed DuSPA algorithm with $\tau = 0.01, \gamma = 5$. It follows from the figure that the outputs of the generators can gradually find their new optimal working points until they reach their limits.

C. Generator Plug-and-Play Analysis

In this section, we report a plug-and-play analysis of the proposed algorithm when certain generator is disconnected from the system and then reconnected to it after a period of time. We consider the same IEEE 14-bus system as before. We show that the proposed algorithm is resilient to generator plug-and-play in the sense that when one generator is disconnected from the system, the power of the generator will be compensated by the corresponding increase of power of other generators so as to meet the load demand and the system can recover quickly once the generator is reconnected to the system. Fig. 3 illustrates the above scenario and the performance of the algorithm under generator plug-and-play. In particular, the power system starts from an optimal point where the power supply meets the demand. At iteration 150, the generator 3 is suddenly disconnected from the system, leading to a shortage of power supply (cf. Fig. 3b). Thus, the power, which is supposed to be produced by the generator

3, now has to be supplied by all the other generators that are still operating within their capacity limits (cf. Fig. 3a). After several iterations, the system gradually reaches to a new equilibrium where the load demand is met again. At iteration 300, the generator 3 is reconnected to the system, resulting in an overshoot of power supply (cf. Fig. 3b). After that, the system gradually goes back to the original equilibrium.

Fig. 3. Illustration of the resilient performance of the algorithm under generator plug-and-play. The generator at Bus 3 is disconnected from the system at iteration 150 and reconnected to the system at iteration 300 onwards. (a) plots the historical evolution of the power output of the generators at Buses 1, 2, 3, 6 and 8 under the proposed algorithm with $\tau = 0.01, \gamma = 5$. (b) plots the evolution of the total power mismatch.

VI. CONCLUSION

In this paper, we proposed a novel distributed algorithm based on dual splitting techniques to solve general composite resource allocation problems which subsumes many real application areas such as smart grid, demand response and network utility maximization. We have showed that, for general convex cost functions, the algorithm is able to find the optimal solution pair that solves the OCEP problem with a *non-ergodic* convergence rate of $O(\frac{1}{k})$ as long as the stepsize is chosen within the specified upperbound. When the cost function is smooth and strongly convex, we have also showed its linear convergence property. In addition, the proposed DuSPA algorithm outperforms the state-of-the-art with proper parameter settings and allows for generator plug-and-play.

It is especially suitable for applications where the proximal operator of the regularization term has a closed form solution, e.g., sparsity-prior, and works very well even for large-scale systems. Since we restrict our attention to undirected fixed networks, it is of great interest to consider more general networks whose topology may be changing over time. This will pose some challenge as the conservation property will no longer hold. Also, since we assume that each agent is willing to cooperate, which is not the case in some situations, it is also of great importance to study the performance of the proposed algorithm under malicious attacks.

APPENDIX A PROOF OF LEMMA 3

Adding (14a) and (14b) we obtain $\mathbf{W}\Delta\mathbf{x}_{k+1} = \tau\mathbf{y}_{k+1} - \tau\mathbf{z}_k$, which can be further rewritten as

$$\mathbf{W}\Delta\mathbf{x}_{k+1} = \tau\mathbf{y}_{k+1} + \tau\Delta\mathbf{z}_{k+1} - \tau\mathbf{z}_{k+1}. \quad (54)$$

Thus, for any $(\mathbf{x}^*, \mathbf{y}^*) \in \mathcal{S}$ and $\mathbf{z}^* = \mathbf{y}^*$, we have

$$\begin{aligned} & \langle \mathbf{W}\Delta\mathbf{x}_{k+1}, \mathbf{x}_{k+1} - \mathbf{x}^* \rangle \\ &= \tau \langle \Delta\mathbf{z}_{k+1}, \mathbf{x}_{k+1} - \mathbf{x}^* \rangle + \langle \tau(\mathbf{y}_{k+1} - \mathbf{y}^*), \mathbf{x}_{k+1} - \mathbf{x}^* \rangle \\ & \quad - \langle \tau(\mathbf{z}_{k+1} - \mathbf{z}^*), \mathbf{x}_{k+1} - \mathbf{x}^* \rangle. \end{aligned} \quad (55)$$

Let $\mathbf{y}_k = (\mathbf{I} - \mathbf{W})\mathbf{y}'_k$ with $\mathbf{y}'_k \in \text{span}^\perp\{\mathbf{1}\}$ (cf. Lemma 2). For the middle term of (55), using (14b) we have

$$\begin{aligned} & \langle \mathbf{y}_{k+1} - \mathbf{y}^*, \mathbf{x}_{k+1} - \mathbf{x}^* \rangle \\ &= \langle \mathbf{y}'_{k+1} - \mathbf{y}^{*'}, (\mathbf{I} - \mathbf{W})(\mathbf{x}_{k+1} - \mathbf{x}^*) \rangle \\ &= -\tau \langle \mathbf{y}'_{k+1} - \mathbf{y}^{*'}, \mathbf{y}_{k+1} - \mathbf{y}_k \rangle \\ &= -\tau \langle \mathbf{y}_{k+1} - \mathbf{y}^*, \mathbf{y}_{k+1} - \mathbf{y}_k \rangle_{\mathbf{Q}^{-1}}, \end{aligned} \quad (56)$$

where, as before, we have used the relation $s = \mathbf{Q}s'$, $\forall s, s' \in \text{span}^\perp\{\mathbf{1}\}$ to obtain the last equality (cf. Eq. (19)).

Recalling that $-\Delta\mathbf{z}_{k+1} + \gamma\hat{\mathbf{x}}_{k+1} - \gamma\nabla f(\mathbf{z}_k) \in \gamma\partial g(\mathbf{z}_{k+1})$ from (14c), together with $\gamma\mathbf{x}^* - \gamma\nabla f(\mathbf{z}^*) \in \gamma\partial g(\mathbf{z}^*)$ from the optimality condition (8c) and the fact that $\mathbf{z}^* = \mathbf{y}^*$, and knowing that ∂g is maximally monotone we have

$$\begin{aligned} & \langle \Delta\mathbf{z}_{k+1} + \gamma(\hat{\mathbf{x}}_{k+1} - \mathbf{x}^*) \\ & \quad - \gamma(\nabla f(\mathbf{z}_k) - \nabla f(\mathbf{z}^*)), \mathbf{z}_{k+1} - \mathbf{z}^* \rangle > 0, \end{aligned} \quad (57)$$

which is equivalent to

$$\begin{aligned} & \gamma \langle \mathbf{x}_{k+1} - \mathbf{x}^*, \mathbf{z}_{k+1} - \mathbf{z}^* \rangle + \gamma \langle \Delta\mathbf{x}_{k+1}, \mathbf{z}_{k+1} - \mathbf{z}^* \rangle \\ & \geq \langle \Delta\mathbf{z}_{k+1}, \mathbf{z}_{k+1} - \mathbf{z}^* \rangle + \gamma \langle \nabla f(\mathbf{z}_k) - \nabla f(\mathbf{z}^*), \mathbf{z}_{k+1} - \mathbf{z}^* \rangle. \end{aligned} \quad (58)$$

Multiplying (58) with $\frac{\tau}{\gamma}$ and combining with (55), (56) yield

$$\begin{aligned} & \langle \mathbf{W}\Delta\mathbf{x}_{k+1}, \mathbf{x}_{k+1} - \mathbf{x}^* \rangle + \tau^2 \langle \mathbf{y}_{k+1} - \mathbf{y}^*, \Delta\mathbf{y}_{k+1} \rangle_{\mathbf{Q}^{-1}} \\ & \quad + \frac{\tau}{\gamma} \langle \Delta\mathbf{z}_{k+1} + \gamma(\nabla f(\mathbf{z}_k) - \nabla f(\mathbf{z}^*)), \mathbf{z}_{k+1} - \mathbf{z}^* \rangle \\ & \quad - \tau \langle \Delta\mathbf{z}_{k+1}, \mathbf{x}_{k+1} - \mathbf{x}^* \rangle - \tau \langle \Delta\mathbf{x}_{k+1}, \mathbf{z}_{k+1} - \mathbf{z}^* \rangle \leq 0, \end{aligned} \quad (59)$$

Since $\gamma < \frac{\lambda_{\min}(\mathbf{W})}{\tau}$ and thus $\mathbf{H} > 0$ (cf. Eq. (27)), the above equation can be rewritten in a compact form as

$$\langle \Delta\mathbf{p}_{k+1}, \mathbf{p}_{k+1} - \mathbf{p}^* \rangle_{\mathbf{H}} \leq -\tau \langle \nabla f(\mathbf{z}_k) - \nabla f(\mathbf{z}^*), \mathbf{z}_{k+1} - \mathbf{z}^* \rangle. \quad (60)$$

Using the following identity in $\mathbf{H}$ -space

$$\begin{aligned} & 2 \langle \mathbf{a} - \mathbf{b}, \mathbf{a} - \mathbf{c} \rangle_{\mathbf{H}} \\ &= \|\mathbf{a} - \mathbf{c}\|_{\mathbf{H}}^2 - \|\mathbf{b} - \mathbf{c}\|_{\mathbf{H}}^2 + \|\mathbf{a} - \mathbf{b}\|_{\mathbf{H}}^2, \forall \mathbf{a}, \mathbf{b} \in \mathcal{H} \end{aligned} \quad (61)$$

for the left-hand side of (60), we complete the proof.

REFERENCES

- [1] J. Xu, S. Zhu, Y. C. Soh, and L. Xie, "A dual splitting algorithm for distributed resource allocation problems," in *Proceedings of IEEE 56th Annual Conference on Decision and Control (CDC)*, 2017, pp. 4674–4679.
- [2] D. K. Molzahn, F. Dörfler, H. Sandberg, S. H. Low, S. Chakrabarti, R. Baldick, and J. Lavaei, "A survey of distributed optimization and control algorithms for electric power systems," *IEEE Transactions on Smart Grid*, 2017.
- [3] L. Xiao, M. Johansson, and S. P. Boyd, "Simultaneous routing and resource allocation via dual decomposition," *IEEE Trans. Commun.*, vol. 52, no. 7, pp. 1136–1144, 2004.
- [4] J. Cortes, S. Martinez, T. Karatas, and F. Bullo, "Coverage control for mobile sensing networks," *IEEE Transactions on Robotics and Automation*, vol. 20, no. 2, pp. 243–255, 2004.
- [5] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, Jan 2011.
- [6] 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.
- [7] R. Olfati-Saber and R. M. Murray, "Consensus problems in networks of agents with switching topology and time-delays," *IEEE Trans. Autom. Control*, vol. 49, no. 9, pp. 1520–1533, 2004.
- [8] L. Xiao and S. Boyd, "Optimal scaling of a gradient method for distributed resource allocation," *Journal of Optimization Theory and Applications*, vol. 129, no. 3, pp. 469–488, 2006.
- [9] T. T. Doan and A. Olshevsky, "On the geometric convergence rate of distributed economic dispatch/demand response in power networks," *arXiv preprint arXiv:1609.06660*, 2016.
- [10] J. Xu, S. Zhu, Y. C. Soh, and L. Xie, "Augmented distributed gradient methods for multi-agent optimization under uncoordinated constant stepsizes," in *Proceedings of 54th IEEE Conference on Decision and Control (CDC)*, 2015, pp. 2055–2060.
- [11] A. Nedich, A. Olshevsky, and W. Shi, "Achieving geometric convergence for distributed optimization over time-varying graphs," *arXiv preprint arXiv:1607.03218*, 2016.
- [12] G. Qu and N. Li, "Harnessing smoothness to accelerate distributed optimization," *IEEE Transactions on Control of Network Systems*, 2017.
- [13] W. Shi, Q. Ling, G. Wu, and W. Yin, "EXTRA: An exact first-order algorithm for decentralized consensus optimization," *SIAM Journal on Optimization*, vol. 25, no. 2, pp. 944–966, 2015.
- [14] J. Xu, S. Zhu, Y. C. Soh, and L. Xie, "A Bregman splitting scheme for distributed optimization over networks," *IEEE Transactions on Automatic Control*, 2018.
- [15] —, "A forward-backward Bregman splitting scheme for regularized distributed optimization problems," in *Proceedings of 55th IEEE Conference on Decision and Control (CDC)*, 2016, pp. 1093–1098.
- [16] A. Beck, A. Nedic, A. Ozdaglar, and M. Teboulle, "An $o(1/k)$ gradient method for network resource allocation problems," *IEEE Transactions on Control of Network Systems*, vol. 1, no. 1, pp. 64–73, 2014.
- [17] 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.
- [18] S. Kar and G. Hug, "Distributed robust economic dispatch in power systems: A consensus + innovations approach," in *2012 IEEE Power and Energy Society General Meeting*, 2012, pp. 1–8.
- [19] 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. 4416–4426, 2013.
- [20] A. Cherukuri and J. Cortés, "Distributed generator coordination for initialization and anytime optimization in economic dispatch," *IEEE Transactions on Control of Network Systems*, vol. 2, no. 3, pp. 226–237, 2015.

- [21] A. Cherukuri and J. Cortes, "Initialization-free distributed coordination for economic dispatch under varying loads and generator commitment," *Automatica*, vol. 74, pp. 183–193, 2016.
- [22] P. Yi, 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, pp. 259–269, 2016.
- [23] X. He, D. W. Ho, T. Huang, J. Yu, H. Abu-Rub, and C. Li, "Second-order continuous-time algorithms for economic power dispatch in smart grids," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2017.
- [24] X. Zeng, P. Yi, Y. Hong, and L. Xie, "Continuous-time distributed algorithms for extended monotropic optimization problems," *arXiv preprint arXiv:1608.01167*, 2016.
- [25] L. Condat, "A primal-dual splitting method for convex optimization involving lipschitzian, proximable and linear composite terms," *J. Optim. Theory Appl.*, vol. 158, no. 2, pp. 460–479, 2013.
- [26] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer Science & Business Media, 2011.
- [27] R. T. Rockafellar, "Monotone operators and the proximal point algorithm," *SIAM J. Control Optim.*, vol. 14, no. 5, pp. 877–898, 1976.
- [28] A. J. Wood and B. F. Wollenberg, *Power Generation, Operation, and Control*. John Wiley & Sons, 2012.
- [29] Z. Wang, F. Liu, S. H. Low, C. Zhao, and S. Mei, "Distributed frequency control with operational constraints, Part I: Per-node power balance," *IEEE Transactions on Smart Grid*, vol. PP, no. 99, pp. 1–1, 2017.
- [30] —, "Distributed frequency control with operational constraints, Part II: Network power balance," *IEEE Transactions on Smart Grid*, vol. PP, no. 99, pp. 1–1, 2017.
- [31] W. Meng and X. Wang, "Distributed energy management in smart grid with wind power and temporally coupled constraints," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 8, pp. 6052–6062, Aug 2017.
- [32] Y. Ma, W. Zhang, W. Liu, and Q. Yang, "Fully distributed social welfare optimization with line flow constraint consideration," *IEEE Transactions on Industrial Informatics*, vol. 11, no. 6, pp. 1532–1541, Dec 2015.
- [33] A. Nedic and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Trans. Autom. Control*, vol. 54, no. 1, pp. 48–61, Jan 2009.
- [34] B. Ghahserifard and J. Cortes, "Distributed continuous-time convex optimization on weight-balanced digraphs," *IEEE Trans. Autom. Control*, vol. 59, no. 3, pp. 781–786, March 2014.
- [35] B. T. Polyak, *Introduction to Optimization*. New York: Optimization Software Inc., 1987.
- [36] B. He and X. Yuan, "Convergence analysis of primal-dual algorithms for a saddle-point problem: From contraction perspective," *SIAM J. Imaging Sci.*, vol. 5, no. 1, pp. 119–149, 2012.
- [37] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Systems & Control Letters*, vol. 53, no. 1, pp. 65–78, 2004.

Jinming Xu received the B.S. degree in mechanical engineering from Shandong University, China, in 2009 and the Ph.D. degree in electrical engineering from Nanyang Technological University (NTU), Singapore, in 2016. He was a research fellow in the School of Electrical and Electronic Engineering at NTU from 2016 to 2017. He is currently a postdoctoral scholar in the Ira A. Fulton Schools of Engineering, Arizona State University. His research interests are primarily in distributed optimization and control, and its applications in large-scale signal

processing, machine learning and networked dynamical systems.

Shanying Zhu (S'12–M'15) received the B.S. degree in information and computing science from North China University of Water Resources and Electric Power, Zhengzhou, China, in 2006, the M.S. degree in applied mathematics from Huazhong University of Science and Technology, Wuhan, China, in 2008, and the Ph.D. degree in control theory and control engineering from Shanghai Jiao Tong University, Shanghai, China, in 2013. From 2013 to 2015, he was a research fellow with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. He joined Shanghai Jiao Tong University, China, in 2015, where he is currently an associate professor in Department of Automation, School of Electronic Information and Electrical Engineering. His research interests focus on multi-agent systems and wireless sensor networks, particularly in coordination control of mobile robots and distributed detection and estimation in sensor networks and their applications in industrial networks.

Yeng Chai Soh received the B.Eng. (Hons. I) degree in electrical and electronic engineering from the University of Canterbury, New Zealand, and the Ph.D. degree in electrical engineering from the University of Newcastle, Australia. He joined the Nanyang Technological University, Singapore, after his PhD study and is currently a professor in the School of Electrical and Electronic Engineering. Dr Soh has served as the Head of the Control and Instrumentation Division, the Associate Dean (Research and Graduate Studies) and the Associate

Dean (Research) at the College of Engineering. Dr Soh has served as panel members of several national grants and scholarships evaluation and awards committees.

Dr Soh's research interests have been in robust control and applications, robust estimation and filtering, optical signal processing, and energy efficient systems. He has published more than 240 refereed journal papers in these areas. His most recent research projects and activities are in sensor networks, sensor fusion, distributed control & optimization, and control & optimization of AC/MV systems.

Lihua Xie (S'91–M'92–SM'97–F'07) received the B.E. and M.E. degrees in electrical engineering from Nanjing University of Science and Technology in 1983 and 1986, respectively, and the Ph.D. degree in electrical engineering from the University of Newcastle, Australia, in 1992. Since 1992, he has been with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, where he is currently a professor and served as the Head of Division of Control and Instrumentation from July 2011 to June 2014. He held teaching

appointments in the Department of Automatic Control, Nanjing University of Science and Technology from 1986 to 1989.

Dr Xie's research interests include robust control and estimation, networked control systems, multi-agent networks, localization and unmanned systems. He is an Editor-in-Chief for Unmanned Systems and an Associate Editor for IEEE Transactions on Network Control Systems. He has served as an editor of IET Book Series in Control and an Associate Editor of a number of journals including IEEE Transactions on Automatic Control, Automatica, IEEE Transactions on Control Systems Technology, and IEEE Transactions on Circuits and Systems-II. He is an elected member of Board of Governors, IEEE Control System Society (Jan 2016–Dec 2018). Dr Xie is a Fellow of IEEE and Fellow of IFAC.