An Augmented Lagrangian Distributed Algorithm for an In-network Optimal Resource Allocation Problem

Solmaz S. Kia (Appeared in the proceeding of ACC 2017)

Abstract— **This paper studies distributed solutions for an optimal resource allocation problem over networked systems with connected graph communication topologies. The problem setting consists of a group of agents in a network cooperatively meeting a demand by supplying a resource whose commitment incurs a cost on them. The objective in the optimal resource allocation problem is to obtain a commitment value for each agent such that the total cost, which is the sum of the costs of the agents, is minimized. In this paper we discuss how using ideas from Augmented Lagrangian method for convex optimization problems with affine constrains, we can arrive at a distributed solutions whose convergence guarantees holds for networks where the local costs are convex. We also show that if the local costs are all strongly convex and their gradients are globally Lipschitz then the convergence guarantees are exponential and the results can be extended to a special class of time-varying network interaction topologies. Simulations illustrate our results.**

I. Introduction

In this paper, we focus on design of a distributed algorithm for a class of in-network optimal resource allocation problem. In-network optimal resource allocation problem appears in many optimal decision making tasks such as economic dispatch over power networks [1]–[3], optimal routing [4]–[7], network resource allocation for wireless systems [8]–[10], economic systems [11], and health care applications [12], [13]. In optimal in-network resource allocation a group of agents work together towards meeting a demand by each committing a local resource. The goal here is to find the optimal commitment value for each agent such that the overall cost of the operation, which is consisted of the sum of the local cost of each agent to provide its commitment, is minimized. Desire for greater autonomy and privacy at agent level and also for avoiding shortcomings such as large data processing and data storage demand at the central node, scalability and, more importantly, the existence of a single failure point for the system have created a demand for decentralized leader less algorithms for cooperative multi-agent systems including optimization algorithms.

In many optimal resource allocation problems, the cost is modeled as a convex function, therefore, the optimal resource allocation problem is a convex optimization problem. In recent years, there has been a renewed attention on decentralized algorithm design for in-network convex optimization problems, and various continuous-time ([14]–[17]) and discrete-time (see e.g., [18]–[22]) are proposed. The underlying principle in design of many of distributed optimization algorithms is to have each agent to maintain a copy of the global optimal decision variable and evolve this local copy using a local iterative algorithm consisted of both local variables and and also the average of local copies of the agent itself and its neighbors. In a network of N agent, for optimal resource allocation problem, such an approach will require each agent to maintain a copy of the global variable which is of size N resulting in computational, communication and storage costs of at least order N per agent. Such a requirement is costly and unnecessary for this problem as the agents only need to obtain their own respective component of the global decision variable. This approach is not suitable for optimal resource allocation problem as here each agent $i \in \{1, \dots, N\}$ is interested in obtaining its own optimal commitment value $\mathsf{x}_i^\star$ in the optimal network variable $\mathbf{x}^\star = (\mathsf{x}_1^\star, \cdots, \mathsf{x}_1^\star)$.

Distributed optimization algorithms targeting the optimal in-network resource allocation problems are presented in [23] in discrete-time form, and in [3], [24]–[26] in continuous-time form. These algorithms require the agents to keep and evolve only their respective component of the global decision variable. However, [3], [23], [24] require the agents to transmit the gradient of local cost functions directly to their neighbors which makes these algorithms less favorable for privacy-sensitive applications. The algorithms in [25], [26] do not have such a requirement. However, the algorithm of [25] comes with higher computational complexity. In comparison to the algorithm we propose in this paper, the algorithm of [26] requires one more variable to be transmitted to the neighbors and also shows less favorable transient response (see Section VI for a numerical example). Moreover, the convergence guarantees of [26] are only for strictly convex local cost functions. In the context of economic dispatch problem, [27]–[29] offer distributed solutions for convex, quadratic local cost functions for the power generators.

Our focus in this paper is to design a distributed algorithm which solves the in-network resource allocation problem with a low local computational complexity per agent. Also, out of concern for the privacy of agents, we are interest in designing an algorithm that does not require the agents to communication any information about their local cost functions to their neighbors. In this paper, we extend our earlier work in [30] which proposes an algorithm with a simple third order dynamics per agent, in which, agent arrive at their own optimal commitment value by only exchanging a logical variable with no need to share their local gradients with their neighbors. Our contribution in this paper includes proposing a novel algorithm which is designed based on ideas from Augmented Lagrangian method in the optimization literature (cf. e.g. [20], [32]) to produce better transient behavior compared to [30]. We show that using the Augmented Lagrangian method, we can extend our rig-

The author is with the Dept. of Mechanical and Aerospace Engineering, University of California Irvine, Irvine, CA 92697, solmaz@uci.edu

orous convergence guarantees from for only strictly convex local costs to convex local cost functions. Our convergence analysis is based on Lyapunov and LaSalle invariant set methods. We also use semistability analysis results to show that our algorithm is guaranteed to converge to a point in the set of optimal decision values when the local costs are convex. In the literature such guarantees are normally for convergence to the set of optimal points as opposed to a point in the set. Finally, we show that when the local cost functions are strongly convex and their local gradients are globally Lipschitz the convergence guarantees of our proposed algorithm over connected graphs is exponential and can also be extended to dynamic graphs. Due to the space limitations, the proofs are omitted and will appear elsewhere.

II. Notations and Definitions

We let $\mathbf{1}_n$ (resp. $\mathbf{0}_n$) denote the vector of n ones (resp. n zeros), and denote by $\mathbf{I}_n$ the $n\times n$ identity matrix. When clear from the context, we do not specify the matrix dimensions. We denote the standard Euclidean norm of vector $\mathbf{x}\in\mathbb{R}^n$ by $\|\mathbf{x}\|=\sqrt{\mathbf{x}^\top\mathbf{x}}$. A differentiable function $f:\mathbb{R}^d\to\mathbb{R}$ is convex over a convex set $C\subseteq\mathbb{R}^d$ iff $(\mathsf{z}-\mathsf{x})^\top(\nabla f(\mathsf{z})-\nabla f(\mathsf{x}))\geq 0$ for all $\mathsf{x},\mathsf{z}\in C$. When this inequality condition is strict, we refer to the function as strictly convex and when we have $(\mathsf{z}-\mathsf{x})^\top(\nabla f(\mathsf{z})-\nabla f(\mathsf{x}))\geq m(\mathsf{z}-\mathsf{x})^\top(\mathsf{z}-\mathsf{x})$ for $m\in\mathbb{R}_{>0}$ and all $\mathsf{x},\mathsf{z}\in C$ we refer to the function f as $m-$strongly convex. Here, $\nabla f(\mathbf{x})$ represents the gradient of f. Function $f:\mathbb{R}^n\to\mathbb{R}^m$ is M-globally Lipschitz, $M\in\mathbb{R}_{>0}$, iff $\|f(\mathsf{z})-f(\mathsf{x})\|\leq M\|\mathsf{z}-\mathsf{x}\|$ for all $\mathsf{x},\mathsf{z}\in\mathbb{R}^n$. In a network of N agents, to distinguish and to emphasize that a variable is local to an agent $i\in\{1,\dots,N\}$, we use superscripts, e.g., $f^i(x^i)$ is the local function of agent i evaluated at its own local state x^i. Moreover, if $v^i\in\mathbb{R}$ is a variable of agent $i\in\{1,\dots,N\}$, the aggregated v^i's of the network is the vector $\mathbf{v}=[v^1,\cdots,v^N]^\top\in\mathbb{R}^N$.

Next, we briefly review some relevant basic concepts from algebraic graph theory following [33]. A weighted *graph*, is a triplet $\mathcal{G}=(\mathcal{V},\mathcal{E},\mathbf{A})$, where $\mathcal{V}=\{1,\dots,N\}$ is the *node set*, $\mathcal{E}\subseteq\mathcal{V}\times\mathcal{V}$ is the *edge set*, and $\mathbf{A}\in\mathbb{R}^{N\times N}$ is a weighted *adjacency* matrix such that $\mathsf{a}_{\mathsf{ij}}>0$ if $(i,j)\in\mathcal{E}$ and $\mathsf{a}_{ij}=0$, otherwise. An edge from i to j, denoted by (i,j), means that agent j can send information to agent i. A graph is *undirected* if $(i,j)\in\mathcal{E}$ anytime $(j,i)\in\mathcal{E}$. An undirected graph is called *connected graph* if $\mathsf{a}_{ij}=\mathsf{a}_{ji}$ for all $i,j\in\{1,\dots,N\}$ and there is path from every node to every other node in the network. The (out-) *Laplacian* matrix of a graph is $\mathbf{L}=\mathrm{Diag}(\mathbf{A}\mathbf{1}_N)-\mathbf{A}$. Based on the structure of $\mathbf{L}$, at least one of the eigenvalues of $\mathbf{L}$ is zero and the rest of them have nonnegative real parts. Moreover, we always have $\mathbf{L}\mathbf{1}_N=\mathbf{0}$. A graph is connected iff $\mathbf{1}_N^\top\mathbf{L}=\mathbf{0}$, and $\mathrm{rank}(\mathbf{L})=N-1$. Therefore, for a connected graph zero is a simple eigenvalue of $\mathbf{L}$. We denote the eigenvalues of $\mathbf{L}$ by $\{\lambda_i\}_{i=1}^N$, where $\lambda_1=0$ and $\lambda_i\leq\lambda_j$, for $i<j$.

III. Problem Statement

We consider the optimal resource allocation problem

$$\mathsf{x}^\star=\arg\min_{\mathbf{x}\in\mathbb{R}^N}\ \sum\nolimits_{i=1}^N f^i(x^i),\quad\text{subject to}\tag{1a}$$

$$x^1+\cdots+x^N=\mathsf{b},\tag{1b}$$

over a network of N agents, communicating with each other over a connected graph topology $\mathcal{G}$. Here, $f^i:\mathbb{R}\to\mathbb{R}$ is the convex and differentiable local cost of agent $i\in\{1,\dots,N\}$, $x^i\in\mathbb{R}$ is its resource, and $\mathsf{b}\in\mathbb{R}$ is the demand that the agents should meet cooperatively (see (1b)).

The objective in our distributed optimal resource allocation problem is to enable each agent $i\in\{1,\dots,N\}$ to obtain, in a distributed manner, $\mathsf{x}_i^\star$, the i^{th} element of $\mathsf{x}^\star=(\mathsf{x}_1^\star,\cdots,\mathsf{x}_N^\star)$ in (1). For agent $i\in\{1,\dots,N\}$, $\mathsf{x}_i^\star$ indicates its resource commitment for minimal collective cost for the network. By distributed we mean that every agent obtains its optimal operating point by only interacting with its neighboring agents on the communication graph $\mathcal{G}$. To this end, we assume that every agent starts a local state $x^i\in\mathbb{R}$ with an initial guess $x^i(0)\in\mathbb{R}$ and evolves it according to

$$\dot{x}^i=c^i(t),\tag{2}$$

until x^i converges to $\mathsf{x}_i^\star$. To respect the requirement for distribute operation, the driving command $c^i(t)$ can only be a function of the local variable and the variables agent i receives from its neighbors. Our objective, then, is to design $c^i(t)$ of each agent $i\in\{1,\dots,N\}$. We make the following assumption about the feasibility of the optimization problem (1).

Assumption 1: The optimization problem (1) has a finite optimum $f^\star=f(\mathsf{x}^\star)$ with a $\mathsf{x}^\star\in X_{\text{fe}}=\{\mathbf{x}\in\mathbb{R}^N\,|\,x^1+\cdots+x^N=\mathsf{b}\}$.

One can use (smooth or non-smooth) penalty function methods to address optimal resource allocation problem subject to bounded resources, i.e., when the optimization problem (1) is constrained by local convex inequality constraints $\mathsf{x}_{\mathsf{l}}^i\leq x^i\leq\mathsf{x}_{\mathsf{u}}^i$, where $\mathsf{x}_{\mathsf{l}}^i\in\mathbb{R}$ and $\mathsf{x}_{\mathsf{u}}^i\in\mathbb{R}$, with $\mathsf{x}_{\mathsf{l}}^i<\mathsf{x}_{\mathsf{u}}^i$, are, respectively the known lower and upper bounds on the resource of each agent $i\in\{1,\dots,N\}$. In penalty function method the inequality constraint is eliminated by adding a weighted penalty function to the cost. Notice that when the inequality constraints are local to each agent, the global cost in the penalty method stays separable, a characteristics that is normally required to develop distributed solutions. A smooth penalty function method whose weight can be selected in a distributed manner is proposed in [30], and can be used here as well to address the optimal resource allocation with bounded resources. For brevity, the details are eliminated.

IV. A central solution for optimal resource allocation problem

In this section, we review a central solution for the optimization problem (1) which provides insights on how to construct our distributed solution.

The KKT optimality conditions give the following necessary and sufficient conditions to characterize the solution set of the convex optimization problem (1).

Lemma 4.1: (Solution set of (1) *(cf. [34] for proof)):* Consider the constrained optimization problem (1). Let $f^i:\mathbb{R}\to\mathbb{R}$, $i\in\{1,\dots,N\}$, be a differentiable and convex function on $\mathbb{R}$. A point $\mathsf{x}^\star\in\mathbb{R}^N$ is a solution of (1) iff there exists a $\mathsf{y}^\star\in\mathbb{R}$, such that

$$\nabla f^i(\mathsf{x}_i^\star)+\mathsf{y}^\star=0,\quad i\in\{1,\dots,N\},\tag{3a}$$

$$\mathsf{x}_1^\star+\cdots+\mathsf{x}_N^\star-\mathsf{b}=0.\tag{3b}$$

When all the local cost functions are strictly convex, the solution $(\mathsf{y}^\star, \mathbf{x}^\star)$ of (3) is unique. □

In general, solving the KKT equation set (3) in an analytical manner is challenging. Therefore, iterative algorithms normally are employed to arrive to points satisfying the KKT conditions starting from an arbitrary initial guesses. Let

$$L(y, \mathbf{x}) = f(\mathbf{x}) + y(x^1 + \cdots + x^N - \mathsf{b}), \tag{4}$$

be the Lagrangian of optimization problem (1). Then, following [35] we can show that when the cost functions are strictly convex, the following saddle-point dynamics (5) converges to the unique solution of (3), i.e.,

$$\dot{y} = \frac{\partial L(y, \mathbf{x})}{\partial y} = x^1 + \cdots + x^N - \mathsf{b}, \tag{5a}$$

$$\dot{x}^i = -\frac{\partial L(y, \mathbf{x})}{\partial x^i} = -\nabla f^i(x^i) - y, \; i \in \{1, \ldots, N\}, \tag{5b}$$

from any initial condition $y(0), x^i(0) \in \mathbb{R}$, generates trajectory $t \mapsto (y(t), \mathbf{x}(t))$ of (5) which converges to the unique solution $(\mathsf{y}^\star, \mathbf{x}^\star)$ of (3). The structure of the saddle point dynamics (5) resembles that of the dual ascent methodology in iterative discrete-time algorithms. It is well known in the discrete-time optimization literature that the convergence properties of dual ascent method can be improved via the Augmented Lagrangian method, e.g., the convergence guarantees can be extended from only for strictly convex to, indeed, convex functions (cf. e.g. [20], [32]). Augmented Lagrangian algorithms start off with casting the optimization problem (1) in equivalent form (for $\rho \in \mathbb{R}_{>0}$)

$$\mathbf{x}^\star = \arg \min_{\mathbf{x} \in \mathbb{R}^N} \sum\nolimits_{i=1}^{N} f^i(x^i) + \frac{\rho}{2}\|x^1 + \cdots + x^N - \mathsf{b}\|^2, \text{ s.t.} \tag{6a}$$

$$x^1 + \cdots + x^N = \mathsf{b}, \tag{6b}$$

whose Lagrangian is

$$L_a(y, \mathbf{x}) = f(\mathbf{x}) + \frac{\rho}{2}\|(\mathbf{1}_N^\top \mathbf{x} - \mathsf{b})\|^2 + y(\mathbf{1}_N^\top \mathbf{x} - \mathsf{b}). \tag{7}$$

Remark that the augmented cost (6a) is still convex and as a result (6) is a convex optimization problem. We can construct a saddle point dynamics below to solve (6)

$$\dot{y} = \frac{\partial L_a(y, \mathbf{x})}{\partial y} = x^1 + \cdots + x^N - \mathsf{b}, \tag{8a}$$

$$\dot{x}^i = -\frac{\partial L_a(y, \mathbf{x})}{\partial x^i} = -\nabla f^i(x^i) - \rho(x^1 + \cdots + x^N - \mathsf{b}) - y, \tag{8b}$$

for $i \in \{1, \ldots, N\}$. Interestingly, the convergence guarantees for (8) holds for convex functions, as well. A simulation study in Section VI (see Fig. **??**) compares the performance of the saddle point dynamical solver (5), which is constructed from the Lagrangian (4) for the original optimization problem (1), with the performance of saddle point dynamical solver (8), which is constructed from the Augmented Lagrangian (7) of the equivalent optimization problem (6). In simulations we have observed that the augmented solver produces a better transient response. This can be attributed to the introduction of negative x terms in the state equation (8b). The details regarding the central solver's convergence analysis are omitted for brevity, however, we will discuss fully the formal evaluation of the effect of the augmented term on the performance of the solver when we study the convergence of our proposed distributed optimization algorithm that solves (1).

The challenge that the Augmented Lagrangian approach presents for distributed algorithm design is that the augmented global cost function (6a) is no longer separable. However, as we discuss next, our approach which makes use of a distributed mechanism to construct the coupling term $x^1 + \cdots + x^N - \mathsf{b}$ in (8) effectively eliminates this concern.

V. Distributed Augmented Lagrangian Algorithm for Optimal Resource Allocation

The following novel algorithm initialized at $x^i(0), y^i(0), v^i(0) \in \mathbb{R}$ such that $\sum_{i=1}^N v^i(0) = 0$, is our solution for problem (1) over a network with connected graph topology $\mathcal{G}$,

$$\dot{v}^i = \sum\nolimits_{j=1}^{N} \mathsf{a}_{ij}(y^i - y^j), \tag{9a}$$

$$\dot{y}^i = (x^i - \bar{\mathsf{b}}^i) - \sum\nolimits_{j=1}^{N} \mathsf{a}_{ij}(y^i - y^j) - v^i, \tag{9b}$$

$$\dot{x}^i = -\nabla f^i(x^i) - \rho\,(x^i - \bar{\mathsf{b}}^i) + \rho\, v^i - y^i, \tag{9c}$$

for $i \in \{1, \ldots, N\}$ and $\sum_{i=1}^N \bar{\mathsf{b}}^i = \mathsf{b}$. Remark that the requirement $\sum_{i=1}^N v^i(0) = 0$ is trivially satisfied if each agent i starts at $v^i(0) = 0$, and $\sum_{i=1}^N \bar{\mathsf{b}}^i = \mathsf{b}$ can be satisfied using $\bar{\mathsf{b}}^i = \mathsf{b}/N$, $i \in \{1, \ldots, N\}$ (another possibility is $\bar{\mathsf{b}}^j = \mathsf{b}$ and $\bar{\mathsf{b}}^i = 0$, $i \in \{1, \ldots, N\} \backslash \{j\}$).

To demonstrate how algorithm (9) is related to the Augmented Lagrangian solver (8), we first write algorithm (9) in its equivalent representation as follows

$$\dot{\mathbf{v}} = \mathbf{L}\mathbf{y}, \tag{10a}$$

$$\dot{\mathbf{y}} = \mathbf{x} - \bar{\mathbf{b}} - \mathbf{L}\mathbf{y} - \mathbf{v}. \tag{10b}$$

$$\dot{x}^i = -\nabla f^i(x^i) - \rho\, \dot{y}^i - y^i - \rho \sum\nolimits_{j=1}^{N} \mathsf{a}_{ij}(y^i - y^j), \tag{10c}$$

for $i \in \{1, \ldots, N\}$. Then, remark that under aforementioned initialization condition, if the execution of (9) over connected graph converges to a point, i.e., as $t \to \infty$ we get $\dot{x}^i = 0$, $\dot{\mathbf{y}} = \mathbf{0}$, $\dot{\mathbf{v}} = 0$ and $t \mapsto (v^i(t), y^i(t), x^i(t))$ converges to $(\underline{v}^i, \underline{y}^i, \underline{x}^i)$, then we can write, for $i \in \{1, \ldots, N\}$,

$$\mathbf{0} = \mathbf{L}\underline{\mathbf{y}} \qquad \Rightarrow \; \underline{\mathbf{y}} = \theta\, \mathbf{1}_N, \quad \theta \in \mathbb{R}, \tag{11a}$$

$$\mathbf{0} = \underline{\mathbf{x}} - \bar{\mathbf{b}} - \mathbf{L}\underline{\mathbf{y}} - \underline{\mathbf{v}} \qquad \Rightarrow \; \underline{x}^1 + \cdots + \underline{x}^N - \mathsf{b} = 0, \tag{11b}$$

$$\begin{cases} 0 = -\nabla f^i(\underline{x}^i) - \underline{y}^i - \rho \sum_{j=1}^N \mathsf{a}_{ij}(\underline{y}^i - \underline{y}^j) \\ \qquad\qquad \Rightarrow \; 0 = -\nabla f^i(\underline{x}^i) - \theta. \end{cases} \tag{11c}$$

The relation in (11a) is deduced from $\mathrm{rank}(\mathbf{L}) = N - 1$ and $\mathbf{L}\mathbf{1}_N = \mathbf{0}$, the relation in (11b) is the result of left multiplying the equality equation to the left of the arrow by $\mathbf{1}_N^\top$ and using the initialization condition along with $\mathbf{1}^\top \mathbf{L} = \mathbf{0}$, and finally the last relation in (11c) is obtained from substituting for $\underline{y}^i$ from the first equation. As a result the ultimate point $(\underline{v}^i, \underline{y}^i, \underline{x}^i) = (\mathsf{x}_i^\star - \bar{\mathsf{b}}^i, \mathsf{y}^\star, \mathsf{x}_i^\star)$, $i \in \{1, \ldots, N\}$, where $(\mathsf{y}^\star, \mathbf{x}^\star)$ is any solution of the KKT conditions (3) of the optimal resource allocation problem (1). This means that under stated initial conditions, if algorithm (9) is convergent, it solves the optimization problem (1) in a distributed manner. In the following, we shall formally study the convergence of the algorithm (9) using Lyapunov stability analysis tool. However, before our analysis we comment on the idea behind the composition of distributed algorithm (9) which is originated from the central Augmented Lagrangian algorithm (8). To observe this relation, let us multiply (10a) and (10b) by $\mathbf{1}^\top$ from left hand side to obtain

$$\sum\nolimits_{i=1}^{N}\dot{v}^i = 0 \;\Rightarrow\; \sum\nolimits_{i=1}^{N}v^i(t) = \sum\nolimits_{i=1}^{N}v^i(0) = 0, \quad (12a)$$

$$\sum\nolimits_{i=1}^{N}\dot{y}^i = x^1 + \cdots + x^N - \mathsf{b}, \quad (12b)$$

which indicates that the dynamics of the sum of y^is duplicates the Lagrange multiplier dynamics (8a) of the central Augmented Lagrangian method. Here, we used $\mathbf{1}_N^\top \mathbf{L} = \mathbf{0}$ for connected graphs. As discussed above, in a convergent (9), ultimately all the y^i converge to the same value indicating that ultimately every agent obtains a scaled local copy of (8a). Moreover, ultimately, the last term in the right hand side of (10c) also disappears and every agent's x^i state dynamics becomes a local copy of (8b) of the Augmented Lagrangian central algorithm.

The preceding discussion sketched the idea behind the construction of the distributed algorithm (9). In the following, we provide a rigorous study of the stability and convergence properties of the algorithm (9) using Lyapunov stability analysis tools. In our proofs, we use the fact that for connected graphs

$$0 < \lambda_2 \mathbf{I}_{N-1} \leq \mathbf{R}^\top \mathbf{L}\mathbf{R} \leq \lambda_N \mathbf{I}_{N-1}, \quad (13)$$

holds, where $\mathbf{R} \in \mathbb{R}^{N\times(N-1)}$ is a matrix that makes $[\frac{1}{\sqrt{N}}\mathbf{1}_N \quad \mathbf{R}] \in \mathbb{R}^{N\times N}$ an orthonormal matrix. Here, recall that λ_2 and λ_N are, respectively, the smallest non-zero and the largest eigenvalues of the Laplacian matrix $\mathbf{L}$. Remark that $\mathbf{R}$ satisfies

$$\mathbf{r}=\frac{\mathbf{1}_N}{\sqrt{N}},\;\; \mathbf{R}\mathbf{R}^\top + \mathbf{r}\mathbf{r}^\top = \mathbf{I}_N,\;\; \mathbf{r}^\top\mathbf{R}=\mathbf{0},\;\; \mathbf{R}^\top\mathbf{R}=\mathbf{I}_{N-1}. \quad (14)$$

Theorem 5.1: (Asymptotic convergence of (9) *over connected graphs):* Let $\mathcal{G}$ be a connected graph, each f^i, $i \in \{1,\dots,N\}$, be convex and differentiable and $\sum_{i=1}^N \bar{\mathsf{b}}^i = \mathsf{b}$ in (9). Then, for each $i \in \{1,\dots,N\}$, starting from $y^i(0), v^i(0), x^i(0) \in \mathbb{R}$ with $\sum_{i=1}^N v^i(0) = 0$, the algorithm (9) over $\mathcal{G}$, for any $\rho \in (0,1)$, makes $t \mapsto (y^i(t), x^i(t))$ to converge asymptotically to $(\mathsf{y}^\star, \mathsf{x}_i^\star)$, where $(\mathsf{y}^\star, \mathbf{x}^\star)$ is a point satisfying the KKT conditions (3) of problem (1)

Sketch of the proof: Let $\mathbf{h}(\boldsymbol{\chi}) = \boldsymbol{\nabla} f(\boldsymbol{\chi}+\mathbf{x}^\star) - \boldsymbol{\nabla} f(\mathbf{x}^\star)$. Then, represent (9) in the equivalent form

$$\dot{u}_1 = 0, \quad u_1(0) = 0, \quad (15a)$$
$$\dot{\mathbf{u}}_{2:N} = \mathbf{R}^\top \mathbf{L}\mathbf{R}\mathbf{z}_{2:N}, \quad (15b)$$
$$\dot{z}_1 = \mathbf{r}^\top \boldsymbol{\chi}, \quad (15c)$$
$$\dot{\mathbf{z}}_{2:N} = \mathbf{R}^\top \boldsymbol{\chi} - \mathbf{R}^\top\mathbf{L}\mathbf{R}\mathbf{z}_{2:N} - \mathbf{u}_{2:N}, \quad (15d)$$
$$\dot{\boldsymbol{\chi}} = -\mathbf{h}(\boldsymbol{\chi}) - \rho\boldsymbol{\chi} + \rho\mathbf{R}\mathbf{u}_{2:N} - \mathbf{r}z_1 - \mathbf{R}\mathbf{z}_{2:N}, \quad (15e)$$

using the change of variables

$$\begin{bmatrix} u_1 \\ \mathbf{u}_{2:N}\end{bmatrix} = \begin{bmatrix}\mathbf{r}^\top \\ \mathbf{R}^\top\end{bmatrix}(\mathbf{v} - (\mathbf{x}^\star - \bar{\mathbf{b}})), \;\; \begin{bmatrix} z_1 \\ \mathbf{z}_{2:N}\end{bmatrix} = \begin{bmatrix}\mathbf{r}^\top \\ \mathbf{R}^\top\end{bmatrix}(\mathbf{y} - \mathsf{y}^\star\mathbf{1}_N),$$
$$\boldsymbol{\chi} = \mathbf{x} - \mathbf{x}^\star, \quad (16)$$

where $(\mathsf{y}^\star, \mathbf{x}^\star)$ is a solution of the KKT conditions (3). Here, we used $\mathbf{r}^\top\mathbf{v} = 0$ which is the result of (12a) together with the given initial conditions. Remark that (15a) corresponds to the constant of motion (12a). To study the stability in the other variables, consider the candidate Lyapunov function

$$V(\mathbf{u}_{2:N}, \mathbf{z}, \boldsymbol{\chi}) = \frac{1}{2}\big(\mathbf{u}_{2:N}^\top(\mathbf{R}^\top\mathbf{L}\mathbf{R})^{-1}\mathbf{u}_{2:N} + z_1^2 + \boldsymbol{\chi}^\top\boldsymbol{\chi} \quad (17)$$
$$+ (1-\rho)\,\mathbf{z}_{2:N}^\top\mathbf{z}_{2:N} + \rho\,(\mathbf{z}_{2:N}+\mathbf{u}_{2:N})^\top(\mathbf{z}_{2:N}+\mathbf{u}_{2:N})\big).$$

The rest of the proof relies on showing that $\dot{V} \leq 0$ along the trajectories of (15b)-(15e). Next, we invoke the invariant set stability results to prove that the trajectories of (15b)-(15e) converge to its set of equilibrium points. After that, we use the semistability results (c.f., [36, Section 4.7]) to show that the convergence guarantee is in fact to a point in this set. To invoke semistability results, we show in the Appendix that all the equilibrium points of (15b)-(15e) are Lyapunov stable. Note that (15b)-(15e) being semistable implies that starting from any initial condition, the trajectories of (15b)-(15e) converge to one of its equilibrium points. Then, given (16) and (15a), we conclude that starting from stated initial conditions in the statement, the trajectories of (9) converge, as $t \to \infty$, to a point where $(\dot{\mathbf{v}} = 0, \dot{\mathbf{y}} = 0, \dot{\mathbf{x}} = 0)$. As, we showed in (11) and the discussion following it, under the stated initial condition, as $t \to \infty$, the limit point (v^i, y^i, x^i), $i \in \{1,\dots,N\}$ that satisfies $(\dot{\mathbf{v}} = 0, \dot{\mathbf{y}} = 0, \dot{\mathbf{x}} = 0)$ in (9) is equal to $(\mathsf{x}_i^\star - \bar{\mathsf{b}}^i, \mathsf{y}^\star, \mathsf{x}_i^\star)$, where $(\mathsf{y}^\star, \mathbf{x}^\star)$ is a solution of the KKT conditions (3) of the optimal resource allocation problem (1). Notice that the point the algorithm converges to is not necessarily the one that we used in (16). ☐

In Theorem (5.1), the local cost functions are assumed to be convex. Next, we show that when the local cost functions are strongly convex and their gradients are globally Lipschitz, algorithm (9) under appropriate initialization converges to the solution of the optimal resource allocation problem (1) exponentially fast. Recall that because the local cost functions are strongly convex, the minimizer of (1) is unique.

Proposition 5.1: (Exponential convergence of (9) *over connected graphs):* Let $\mathcal{G}$ be a connected graph and $\sum_{i=1}^N \bar{\mathsf{b}}^i = \mathsf{b}$ in (9). Additionally, assume that each f^i, $i \in \{1,\dots,N\}$, is differentiable and m^i-strongly convex and has M^i-Lipschitz gradient where $m^i, M^i \in \mathbb{R}_{>0}$. Then, for each $i \in \{1,\dots,N\}$, starting from $y^i(0), v^i(0), x^i(0) \in \mathbb{R}$ with $\sum_{i=1}^N v^i(0) = 0$, the algorithm (9) over $\mathcal{G}$, for any $\rho \in (0,1)$, makes $t \mapsto (y^i(t), x^i(t))$ to converge exponentially fast to $(\mathsf{y}^\star, \mathsf{x}_i^\star)$ as $t \to \infty$, where, $(\mathsf{y}^\star, \mathbf{x}^\star)$ is the unique solution of the KKT conditions (3) of problem (1).

The extension of the results in Proposition 5.1 to the time-varying connected graph topologies $\mathcal{G}(t)$ whose adjacency matrices are uniformly bounded and piecewise constant is immediate. For such graphs, with $\mathcal{P}$ being the index set of all possible realizations of $\mathcal{G}(t)$, we can write

$$0 < (\lambda_2)_{\min}\mathbf{I}_{N-1} \leq \mathbf{R}^\top\mathbf{L}_p\mathbf{R} \leq (\lambda_N)_{\max}\mathbf{I}_{N-1}. \quad (18)$$

where $(\lambda_2)_{\min} = \min\limits_{p\in\mathcal{P}}\{\lambda_2(\mathbf{L}_p)\}$ and $(\lambda_2)_{\max} = \max\limits_{p\in\mathcal{P}}\{\lambda_2(\mathbf{L}_p)\}$. Then, because the proof of Proposition 5.1 relies on a Lyapunov function that has no dependency on the systems parameters and its derivative is negative definite with a quadratic upper bound in which λ_2 is replaced by $(\lambda_2)_{\min}$, the proof can be deduced from [37, Theorem 4.10]. The proof details are omitted for brevity.

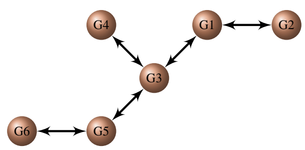


Fig. 1: The graph used in the simulation (adjacency weights are 1).

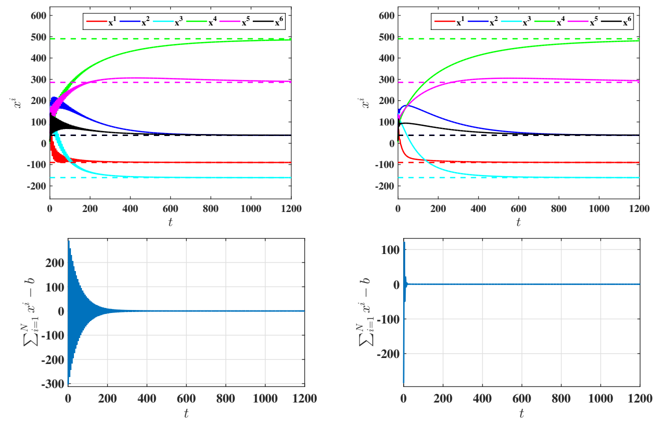
Lemma 5.1 (Convergence of (9) over dynamically changing connected graphs): *Let $\mathcal{G}$ be a time-varying connected graph at all times whose adjacency matrix is uniformly bounded and piecewise constant. Moreover, let* $\sum_{i=1}^{N} \bar{\mathsf{b}}^i = \mathsf{b}$ *in* (9). *Additionally, assume that each f^i, $i \in \{1,\dots,N\}$, is differentiable and m^i-strongly convex and has M^i-Lipschitz gradient where $m^i, M^i \in \mathbb{R}_{>0}$. Then, for each $i \in \{1,\dots,N\}$, starting from $y^i(0), v^i(0), x^i(0) \in \mathbb{R}$ with $\sum_{i=1}^{N} v^i(0) = 0$, the algorithm* (9)*, over $\mathcal{G}$ for any $\rho \in (0,1)$ makes $t \mapsto (y^i(t), x^i(t))$ to converge exponentially fast to $(\mathsf{y}^\star, \mathsf{x}_i^\star)$ as $\to \infty$, where $(\mathsf{y}^\star, \mathbf{x}^\star)$ is the unique solution of the KKT conditions* (3) *of problem* (1).

VI. Simulations

We consider the numerical example of [30] for an economic dispatch problem for a group of 6 generators interacting over a connected graph shown in Fig. 1. The local cost function for each agent is given by $f^i(x^i) = a^i x^{i2} + b^i x^i + c^i$, where $x^i \in \mathbb{R}$ represents the power generated by agent i. Here, a^i, b^i and c^i are constant coefficients with the same values listed in [30, Section 6], selected according to the IEEE 118-bus test model's generators located at buses $(4, 10, 18, 26, 54, 69)$. We assume that the demand load is 600 MW. Fig. 2 compares the response of the distributed solver (9) with augmented term with weight $\rho = 0.5$ vs. when it is implemented with no augmentation term, i.e., when $\rho = 0$. As plots in Fig. 2 show, the Augmented Lagrangian solver is able to remove the undesirable oscillatory behavior from the transient response and to produce faster convergence to the constraint manifold. Similar trend can be observed in responses of the central saddle point dynamical solver (5) and the central Augmented Lagrangian saddle point dynamical solver (8) (due to space limitations the plots are not included). Fig.3 demonstrates the performance of the algorithm in [26].

VII. Conclusions

In this paper we considered an optimal resource allocation problem over a network of N agents communicating with each other over a connected graph. We presented a novel distributed algorithm to solve this problem. To improve the performance of our dynamical solver we invoked the method of Augmented Lagrangian from the convex optimization literature. We showed that using this method the response of the algorithm can be improved and also, the rigorous convergence guarantees for the algorithm can be extended from only for strictly convex local cost functions to convex local cost functions. In the case of convex local cost functions, it is possible that the optimal resource allocation problem to have infinite number of minimizers. We used results from semistability analysis to show that the convergence of our algorithm is indeed to a point in the set of minimizers of the optimal resource allocation problem. When the local cost functions are strongly convex and their gradients are globally Lipschitz, we showed that our algorithm has exponential convergence guarantees. Our future work includes the event-triggered communication implementation of our proposed algorithm, characterization of its privacy preservation properties and rigorous and details analysis of bounded resource optimal resource allocation problem.



(a) trajectories and constraint violation of (9), $\rho = 0$ (b) trajectories and constraint violation of (9), $\rho = 0.5$

Fig. 2: Comparison between the performance of the distributed solver (9) when it is implemented with $\rho = 0$ and when it is implemented with $\rho = 0.5$, over the connected graph depicted in Fig. 1. Both simulations are started from same initial conditions. The colored curved plots depict the time history of the decision variable of each agent. Horizontal lines depict the centralized solution obtained using Matlab's constraint optimization solver 'fmincon'.

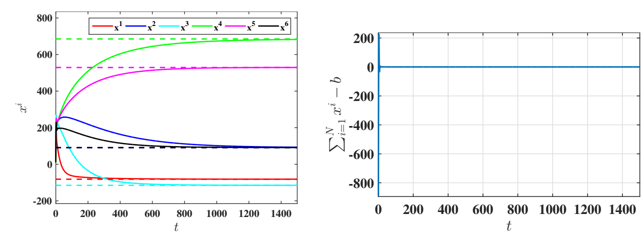


Fig. 3: The colored curved plots depict the time history of the decision variable of each agent when they implement algorithm of [26] over the connected graph depicted in Fig. 1. Horizontal lines depict the centralized solution obtained using Matlab's constraint optimization solver 'fmincon'. Compared to (9), this algorithm demonstrates an oscillatory transient response and lower convergence rate for the demand equation.

Appendix

Complementary part to the proof of Theorem 5.1

We show that all the equilibrium points of (15b)-(15e), i.e., the members of

$$\mathcal{O} = \Big\{ (\underline{\mathbf{u}}_{2:N}, \underline{z}_1, \underline{\mathbf{z}}_{2:N}, \underline{\boldsymbol{\chi}}) \in \mathbb{R}^{N-1} \times \mathbb{R} \times \mathbb{R}^{N-1} \times \mathbb{R}^N \big| \\ \underline{\mathbf{z}}_{2:N} = \mathbf{0},\ \mathsf{r}^\top \underline{\boldsymbol{\chi}} = 0,\ \mathsf{R}^\top \underline{\boldsymbol{\chi}} - \underline{\mathbf{u}}_{2:N} = \mathbf{0},\ \mathbf{h}(\underline{\boldsymbol{\chi}}) + \mathsf{r} \underline{z}_1 = \mathbf{0} \Big\},$$

are all Lyapunov stable. To study the stability of any equilibrium point $(\underline{\mathbf{u}}_{2:N}, \underline{z}_1, \underline{\mathbf{z}}_{2:N}, \underline{\boldsymbol{\chi}}) \in \mathcal{O}$, we transfer that equilibrium point to the origin using $\mathbf{p}=\mathbf{u}_{2:N}-\underline{\mathbf{u}}_{2:N}$, $\boldsymbol{\aleph}=\boldsymbol{\chi}-\underline{\boldsymbol{\chi}}$, $q_1=z_1-\underline{z}_1$, $\mathbf{q}_{2:N}=\mathbf{z}_{2:N}-\underline{\mathbf{z}}_{2:N}$, where $\mathbf{q}=(q_1,\mathbf{q}_{2:N})$, and write (15b)-(15e) in the new coordinate as

$$\dot{\mathbf{p}} = \mathsf{R}^\top\mathsf{L}\mathsf{R}\mathbf{p}_{2:N}, \tag{19a}$$

$$\dot{q}_1 = \mathsf{r}^\top\boldsymbol{\aleph}, \tag{19b}$$

$$\dot{\mathbf{q}}_{2:N} = \mathsf{R}^\top\boldsymbol{\aleph}-\mathsf{R}^\top\mathsf{L}\mathsf{R}\mathbf{q}_{2:N}-\mathbf{p}, \tag{19c}$$

$$\dot{\boldsymbol{\aleph}} = -\tilde{\mathbf{h}}(\boldsymbol{\aleph})-\rho\,\boldsymbol{\aleph}+\rho\,\mathsf{R}\mathbf{p}-\mathsf{r}q_1-\mathsf{R}\mathbf{q}_{2:N}, \tag{19d}$$

where

$$\tilde{\mathbf{h}}(\boldsymbol{\aleph}) = \boldsymbol{\nabla}f(\boldsymbol{\aleph}+\underline{\boldsymbol{\chi}}+\mathsf{x}^\star)-\boldsymbol{\nabla}f(\underline{\boldsymbol{\chi}}+\mathsf{x}^\star). \tag{20}$$

To study the stability of the origin in (19), we use the Lyapunov candidate function (17) in which $(\mathbf{u}_{2:N},\mathbf{z},\boldsymbol{\chi})$ is replaced, respectively, with $(\mathbf{p},\mathbf{q},\boldsymbol{\aleph})$. Taking the derivative of this V along the trajectories of (19), gives

$$\dot{V} = -(1-\rho)\,\mathbf{q}_{2:N}\mathsf{R}^\top\mathsf{L}\mathsf{R}\mathbf{q}_{2:N}-\boldsymbol{\aleph}^\top\tilde{\mathbf{h}}(\boldsymbol{\aleph})-\rho\,\boldsymbol{\aleph}^\top\mathsf{r}\mathsf{r}^\top\boldsymbol{\aleph} - \rho\,(\mathsf{R}^\top\boldsymbol{\aleph}-\mathbf{p})^\top(\mathsf{R}^\top\boldsymbol{\aleph}-\mathbf{p}).$$

Convexity of local cost functions gives $\aleph^i(\boldsymbol{\nabla}f^i(\aleph^i+\underline{\chi}^i+\mathsf{x}_i^\star)-\boldsymbol{\nabla}f^i(\underline{\chi}^i+\mathsf{x}_i^\star)))=(\aleph^i+\underline{\chi}^i+\mathsf{x}_i^\star-(\underline{\chi}^i+\mathsf{x}_i^\star))(\boldsymbol{\nabla}f^i(\aleph^i+\underline{\chi}^i+\mathsf{x}_i^\star)-\boldsymbol{\nabla}f^i(\underline{\chi}^i+\mathsf{x}_i^\star)))\geq 0$, $i\in\{1,\dots,N\}$ for all $t\in\mathbb{R}_{\geq 0}$, which then gives $\boldsymbol{\aleph}^\top\tilde{\mathbf{h}}(\boldsymbol{\aleph})\geq 0$ for all $\boldsymbol{\aleph}\in\mathbb{R}^N$ (recall (20)). Moreover, because the network topology is a connected graph we have $\mathsf{R}^\top\mathsf{L}\mathsf{R}>0$. Therefore, we conclude that along the trajectories of (19) we have $\dot{V}\leq 0$. Therefore, any $(\underline{\mathbf{u}}_{2:N},\underline{z}_1,\underline{\mathbf{z}}_{2:N},\underline{\boldsymbol{\chi}})\in\mathcal{O}$ is a Lyapunov stable equilibrium point. □

REFERENCES

[1] A. J. Wood, F. Wollenberg, and G. B. Sheble, *Power Generation, Operation and Control*. New York: John Wiley, 3rd ed., 2013.

[2] B. H. Chowdhury and S. Rahman, "A review of recent advances in economic dispatch," *IEEE Transactions on Power Systems*, vol. 5, pp. 1248–1259, November 1990.

[3] A. Cherukuri and J. Cortés, "Initiinitialization-freealization-free distributed coordination for economic dispatch under varying loads and generator commitment," *Automatica*, vol. 74, pp. 183–193, 2016.

[4] T. E. Stern, "A class of decentralized routing algorithms using relaxation," *IEEE Transactions on Communications*, vol. COM-25, no. 10, pp. 1092–1102, 1977.

[5] R. G. Gallager, "A minimum delay routing algorithm using distributed computation," *IEEE Transactions on Communications*, vol. COM-25, no. 1, pp. 73–85, 1977.

[6] L. Xiao, M. Johansson, and S. P. Boyd, "Simultaneous routing and resource allocation via dual decomposition," *IEEE Transactions on Communications*, vol. 52, no. 7, pp. 1136–1144, 2004.

[7] R. Madan and S. Lall, "Distributed algorithms for maximum lifetime routing in wireless sensor networks," *IEEE Transactions on Wireless Communications*, vol. 5, no. 8, pp. 2185–2193, 2006.

[8] D. Feijer and F. Paganini, "Stability of primal-dual gradient dynamics and applications to network optimization," *Automatica*, vol. 46, pp. 1974–1981, 2010.

[9] J. Chen and V. K. N. Lau, "Convergence analysis of saddle point problems in time varying wireless systems – control theoretical approach," *IEEE Transactions on Signal Processing*, vol. 60, no. 1, pp. 443–452, 2012.

[10] A. Ferragut and F. Paganini, "Network resource allocation for users with multiple connections: fairness and stability," *IEEE/ACM Transactions on Networking*, vol. 22, no. 2, pp. 349–362, 2014.

[11] G. M. Heal, "Planning without prices," *Review of Economic Studies*, vol. 36, pp. 347–362, 1969.

[12] A. A. Stinnett and A. D. Paltiel, "Mathematical programming for the efficient allocation of health care resource," vol. 15, pp. 641–653, 1996.

[13] S. Weng, T. Wu, G. T. Mackulak, and W. A. Verdini, "A multi-tool integrated methodology for distributed resource allocation in healthcare," vol. 11, no. 4, pp. 428–452, 2012.

[14] J. Wang and N. Elia, "A control perspective for centralized and distributed convex optimization," in *IEEE Int. Conf. on Decision and Control*, (Florida, USA), December 2011.

[15] F. Zanella, D. Varagnolo, A. Cenedese, G. Pillonetto, and L. Schenato, "Newton-Raphson consensus for distributed convex optimization," in *IEEE Int. Conf. on Decision and Control*, (Florida, USA), pp. 5917–5922, December 2011.

[16] J. Lu and C. Y. Tang, "Zero-gradient-sum algorithms for distributed convex optimization: The continuous-time case," *IEEE Transactions on Automatic Control*, vol. 57, no. 9, pp. 2348–2354, 2012.

[17] S. S. Kia, J. Cortés, and S. Martínez, "Distributed convex optimization via continuous-time coordination algorithms with discrete-time communication," *Automatica*, vol. 55, pp. 254–264, 2015.

[18] A. Nedić and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Transactions on Automatic Control*, vol. 54, no. 1, pp. 48–61, 2009.

[19] B. Johansson, M. Rabi, and M. Johansson, "A randomized incremental subgradient method for distributed optimization in networked systems," *SIAM Journal on Optimization*, vol. 20, no. 3, pp. 1157–1170, 2009.

[20] 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, 2010.

[21] M. Zhu and S. Martínez, "On distributed convex optimization under inequality and equality constraints," *IEEE Transactions on Automatic Control*, vol. 57, no. 1, pp. 151–164, 2012.

[22] J. Duchi, A. Agarwal, and M. Wainwright, "Dual averaging for distributed optimization: Convergence analysis and network scaling," *IEEE Transactions on Automatic Control*, vol. 57, no. 3, pp. 592–606, 2012.

[23] 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.

[24] 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.

[25] S. S. Kia, "A distributed dynamical solver for an optimal resource allocation problem over networked systems," in *IEEE Int. Conf. on Decision and Control*, (Osaka, Japan), Dec. 2015.

[26] P. Yi, Y. Hong, and F. Liu, "Initialization-free distributed algorithms for optimal resource allocation with feasibility constraints and its application to economic dispatch of power systems," *Automatica*, vol. 74, pp. 259–269, 2016.

[27] Z. Zhang and M. 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.

[28] S. Kar and G. Hug, "Distributed robust economic dispatch in power systems: A consensus + innovations approach," in *Power & Energy Society General Meeting*, (San Diego, CA), pp. 1–8, July 2012.

[29] A. D. Dominguez-Garcia, S. T. Cady, and C. N. Hadjicostis, "Decentralized optimal dispatch of distributed energy resources," in *IEEE Int. Conf. on Decision and Control*, (Hawaii, USA), pp. 3688–3693, Dec. 2012.

[30] S. S. Kia, "Distributed optimal resource allocation over networked systems and use of an epsilon-exact penalty function," in *IFAC Symposium on Large Scale Complex Systems*, vol. 49, (Riverside, CA), pp. 13–18, May 2016.

[31] M. R. Hestenes, "Multiplier and gradient methods," *Journal of Optimization Theory & Applications*, vol. 4, pp. 302–320, 1969.

[32] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*. Athena Scientific, 1997.

[33] F. Bullo, J. Cortés, and S. Martínez, *Distributed Control of Robotic Networks*. Applied Mathematics Series, Princeton University Press, 2009.

[34] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[35] K. J. Arrow, L. Hurwicz, and H. Uzawa, *Studies in linear and nonlinear programming*. Stanford University Press, 1958.

[36] W. Haddad and V. Chellaboina, *Nonlinear Dynamical Systems and Control*. Princeton, NJ: Princeton University Press, 2008.

[37] H. Khalil, *Nonlinear Systems*. Prentice Hall, 2002.