



Cluster-based distributed augmented Lagrangian algorithm for a class of constrained convex optimization problems[☆]

Hossein Moradian*, Solmaz S. Kia

Department of Mechanical and Aerospace Engineering, University of California, Irvine, United States of America



ARTICLE INFO

Article history:

Received 27 February 2020
Received in revised form 6 January 2021
Accepted 13 February 2021
Available online 13 April 2021

Keywords:

Distributed constrained convex optimization
Augmented Lagrangian
Primal-dual solutions
Optimal resource allocation
Penalty function methods

ABSTRACT

We propose a distributed solution for a constrained convex optimization problem over a network of clustered agents each consisted of a set of subagents. The communication range of the clustered agents is such that they can form a connected undirected graph topology. The total cost in this optimization problem is the sum of the local convex costs of the subagents of each cluster. We seek a minimizer of this cost subject to a set of affine equality constraints, and a set of affine inequality constraints specifying the bounds on the decision variables if such bounds exist. We design our distributed algorithm in a cluster-based framework which results in a significant reduction in communication and computation costs. Our proposed distributed solution is a novel continuous-time algorithm that is linked to the augmented Lagrangian approach. It converges asymptotically when the local cost functions are convex and exponentially when they are strongly convex and have Lipschitz gradients. Moreover, we use an ϵ -exact penalty function to address the inequality constraints and derive an explicit lower bound on the penalty function weight to guarantee convergence to ϵ -neighborhood of the global minimum value of the cost. A numerical example demonstrates our results.

© 2021 Elsevier Ltd. All rights reserved.

1. Introduction

We consider a group of N clustered agents $\mathcal{V} = \{1, \dots, N\}$ with communication and computation capabilities, whose communication range is such that they can form a connected undirected graph topology, see Fig. 1. These agents aim to solve, in a distributed manner, the optimization problem

$$\mathbf{x}^* = \arg \min_{\mathbf{x} \in \mathbb{R}^m} \sum_{i=1}^N f^i(\mathbf{x}^i), \text{ subject to} \quad (1a)$$

$$[\mathbf{w}^1]_j \mathbf{x}^1 + \dots + [\mathbf{w}^N]_j \mathbf{x}^N - b_j = 0, \quad j \in \{1, \dots, p\}, \quad (1b)$$

$$\underline{x}_l^i \leq x_l^i, \quad l \in \underline{\mathcal{B}}^i \subseteq \{1, \dots, n^i\}, \quad i \in \mathcal{V}, \quad (1c)$$

$$\bar{x}_l^i \leq \bar{x}_l^i, \quad l \in \bar{\mathcal{B}}^i \subseteq \{1, \dots, n^i\}, \quad i \in \mathcal{V}, \quad (1d)$$

where $f^i(\mathbf{x}^i) = \sum_{l=1}^{n^i} f_l^i(x_l^i)$. In this setting, each agent $i \in \mathcal{V}$ is a cluster of local 'subagents' $l \in \{1, \dots, n^i\}$ whose decision

variable is $\mathbf{x}^i = [x_1^i, \dots, x_{n^i}^i]^T \in \mathbb{R}^{n^i}$. The weighting factor matrix $\mathbf{w}^i \in \mathbb{R}^{p \times n^i}$ of each agent $i \in \mathcal{V}$ is only known to the agent i itself. Moreover, $\underline{x}_l^i, \bar{x}_l^i \in \mathbb{R}$, with $\underline{x}_l^i < \bar{x}_l^i$, are respectively the lower and upper bounds on the l th decision variable of agent $i \in \mathcal{V}$, if such a bound exists. In a distributed solution, each agent $i \in \mathcal{V}$ should obtain its respective component of $\mathbf{x}^* = [\mathbf{x}^{1*T}, \dots, \mathbf{x}^{N*T}]^T$ by interacting only with the agents that are in its communication range. Problem (1) explicitly or implicitly, captures various in-network optimization problems. One example is the optimal in-network resource allocation, which appears in many optimal decision making tasks such as economic dispatch over power networks (Cherukuri & Cortés, 2016; Wood, Wollenberg, & Sheble, 2013), optimal routing (Madan & Lall, 2006; Xiao, Johansson, & Boyd, 2004) and network resource allocation for wireless systems (Chen & Lau, 2012; Ferragut & Paganini, 2014). In such problems, a group of agents with limited resources, e.g., a group of generators in a power network, add up their local resources to meet a demand in a way that the overall cost is optimum for the entire network. Another family of problems that can be modeled as (1) is the in-network model predictive control over a finite horizon for a group of agents with linear dynamics (Alghunaim, Yuan, & Sayed, 2018; Rostami, Costantini, & Görges, 2017).

In recent years, there has been a surge in the design of distributed algorithms for large-scale in-network optimization problems. The major developments have been in the unconstrained

* This work was supported by NSF, United States of America, CAREER award ECCS-1653838. A preliminary version of this paper is presented in (Kia, 2017a). The material in this paper was partially presented at the 2017 American Control Conference, May 24–26, 2017, Seattle, WA, USA. This paper was recommended for publication in revised form by Associate Editor Luca Schenato under the direction of Editor Christos G. Cassandras.

Corresponding author.

E-mail addresses: hmoradian@uci.edu (H. Moradian), solmaz@uci.edu (S.S. Kia).

convex optimization setting where the global cost is the sum of local costs of the agents (see e.g. Boyd, Parikh, Chu, Peleato, & Eckstein, 2010; Duchi, Agarwal, & Wainwright, 2012 for algorithms in discrete-time, and Kia, 2017a; Kia, Cortés, & Martínez, 2014; Varagnolo, Zanella, Cenedese, Pillonetto, & Schenato, 2015; Wang & Elia, 2011 for algorithms in continuous-time). In-network constrained convex optimization problems have also been studied in the literature. For example, in the context of the power generator economic dispatch problem, Dominguez-Garcia, Cady, and Hadjicostis (2012), Kar and Hug (2012) and Zhang and Chow (2012) offer distributed solutions that solve a special case of (1) with local quadratic costs subject to bounded decision variables and a single demand equation, $p = 1$ and $w^i = 1$ for $i \in \mathcal{V}$. Distributed algorithm design for special cases of (1) with non-quadratic costs is presented in Alghunaim et al. (2018), Xiao and Boyd (2006) and Zhang and Zavlanos (2018) in discrete-time form, and Cherukuri and Cortés (2015, 2016), Ding and Jovanovic (2018), Kia (2017b) and Yi, Hong, and Liu (2016) in continuous-time form. Except for Zhang and Zavlanos (2018), all these algorithms consider the case that the local decision variable of each agent $i \in \mathcal{V}$ is a scalar. Moreover, with the exception of Alghunaim et al. (2018), Kia (2017b) and Zhang and Zavlanos (2018), these algorithms only solve (1) when the equality constraint is the unweighted sum of local decision variables, i.e., $p = 1$ and $w^i = 1$ for $i \in \mathcal{V}$. Also, only Cherukuri and Cortés (2015, 2016) consider local inequality constraints, which are in the form of local box inequality constraints on all the decision variables of the problem. Lastly, the algorithms in Cherukuri and Cortés (2015, 2016) and Xiao and Boyd (2006) require the agents to communicate the gradient of their local cost functions to their neighbors. Such a requirement can be of concern for privacy-sensitive applications.

In this paper, we propose a novel distributed algorithm to solve the optimization problem (1). We start by considering the case that $\underline{\mathcal{B}}^i = \bar{\mathcal{B}}^i = \{\}$ for $i \in \mathcal{V}$, i.e., when there is no inequality constraint. For this problem, we propose a continuous-time distributed primal-dual algorithm. To induce robustness and also to yield convergence without strict convexity of the local cost functions, we adapt an augmented Lagrangian framework (Bertsekas & Tsitsiklis, 1997). The augmented Lagrangian method has been used in Jakovetic, Moura, and Xavier (2015), Vaquero and Cortes (2018), and Zhang and Zavlanos (2018) to improve the transient response of the distributed algorithms for, respectively, an unconstrained convex optimization, an online optimization, and a discrete-time constrained optimization problems. Different than the customary practice of using a common augmented Lagrangian penalty parameter as in Jakovetic et al. (2015), Vaquero and Cortes (2018) and Zhang and Zavlanos (2018), in our design to reduce the coordination overhead among the agents we allow each agent to choose its own penalty parameter locally. The structure of our distributed solution is inspired by the primal-dual centralized solution of Arrow, Hurwicz, and Uzawa (1958) (see (6)), where the coupling in the differential solver is in the dual state dynamics. In decentralized primal-dual algorithms, e.g. Ding, Hu, Dhingra, and Jovanovic (2018), Ding and Jovanovic (2018) and Kia (2016), the adopted practice is to give every agent a copy of the dual variables and use a consensus mechanism to make the agents arrive eventually at the same dual variable. We follow the same approach but in our design, we pay particular attention to computation and communication resource management by adopting a cluster-based approach. First, we consider the sparsity in the equality constraints and give only a copy of a dual variable to an agent if a decision variable of that agent is involved in the equality constraint corresponding to that dual variable. Then, only the cluster of the agents that have a copy of the dual variable need to form a connected graph and use a consensus mechanism to arrive at agreement on their

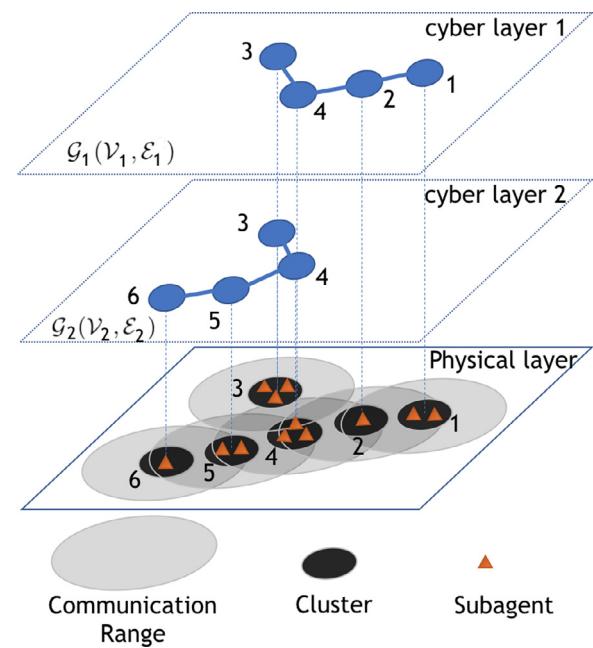


Fig. 1. A group of clustered agents (generators) with undirected connected graph topology aim to solve $\mathbf{x}^* = \arg \min_{\mathbf{x} \in \mathbb{R}^{12}} \sum_{i=1}^6 f^i(\mathbf{x}^i)$, subject to $[1 \ 1]\mathbf{x}^1 + \mathbf{x}^2 + [0.5 \ 0.5 \ 0.5]\mathbf{x}^3 + [1 \ 1]\mathbf{x}^4 = 450$, $[0.5 \ 0.5 \ 0.5]\mathbf{x}^3 + [1 \ 1]\mathbf{x}^5 + \mathbf{x}^6 = 700$, and $\underline{x}_i^l \leq x_i^l \leq \bar{x}_i^l$, $i \in \mathbb{Z}_1^6$, $l \in \mathbb{Z}_1^{n^i}$ in a distributed manner. Here, $f^i(\mathbf{x}^i) = \sum_{l=1}^{n^i} f_l^i(x_l^i)$, where $f_l^i(x_l^i) = \alpha_l^i x_l^{i,2} + \beta_l^i x_l^i + \gamma_l^i$. In the physical layer plot, a cluster agent can communicate with another cluster if it is inside the other cluster's communication disk. To solve this optimal resource allocation problem in a distributed manner, we form subgraphs $\mathcal{G}_1(\mathcal{V}_1, \mathcal{E}_1)$ and $\mathcal{G}_2(\mathcal{V}_2, \mathcal{E}_2)$, which are associated, respectively, with the first and the second equality constraints. Here, $\mathcal{V}_1 = \{1, 2, 3, 4\}$ and $\mathcal{V}_2 = \{3, 4, 5, 6\}$. Agent 4 acts as a connectivity helper node in \mathcal{G}_2 . A solution to this problem using our proposed algorithm is given in Moradian and Kia (2020).

dual variable, see Fig. 1. Next, in our design, we only assign a single copy of the dual variable to an agent i regardless of how many subagents it has. We note that if we use the algorithms in Alghunaim et al. (2018), Cherukuri and Cortés (2015, 2016), Ding and Jovanovic (2018), Kia (2017b), Xiao and Boyd (2006), Yi et al. (2016) and Zhang and Zavlanos (2018) to solve problems where $\mathbf{x}^i \in \mathbb{R}^{n^i}$ of an agent $i \in \mathcal{V}$ is a vector ($n^i > 1$), we need to treat each component of the i as an agent and assign a copy of a dual variable to it. Such a treatment increases the local storage, computation and communication costs of agent i . Our convergence analysis, given in Moradian and Kia (2020), is based on the Lyapunov and the LaSalle invariant set methods, and also the semistability analysis (Haddad & Chellaboina, 2008) 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. 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.

To address scenarios where all or some of the decision variables are bounded in (1), we use a variation of exact penalty function method (Bertsekas, 1975), called ϵ -exact penalty function method (Pinar & Zenios, 1994). Unlike the exact penalty method, this method uses a smooth differentiable penalty function to converge to the ϵ -neighborhood of the global minimum value of the cost. The advantage of exact penalty function methods is in the possibility of using a finite penalty weight to arrive at a practical and numerically well-posed optimization solution. However, as shown in Bertsekas (1975) and Pinar and Zenios (1994), the

penalty function weight is lower bounded by the bounds on the Lagrange multipliers. Since generally, the Lagrange multipliers are unknown, the bound on the penalty function weight is not known either. Many literature that use penalty function methods on distributed optimization framework generally state that a large enough value for the weight is used (Wei, Wang, Li, & Mei, 2017; Zholbarysov, Fooladivanda, & Domínguez-García, 2019), with no guarantees on the feasibility of their choice. Cherukuri and Cortés (2015, Lemma 5.1), Kia (2016, Proposition 4), Mangasarian (1985), and Richter, Morari, and Jones (2011), are among few results in literature that address the problem of establishing an exact upper-bound on the size of the Lagrange multipliers, which can be used to obtain a lower bound on the size of the valid penalty function weight. However, Mangasarian (1985) considers problems with inequality constraints only, while Cherukuri and Cortés (2015, Lemma 5.1), and Kia (2016, Proposition 4) are developed for the resource allocation problem described by (1) when there exists only one equality constraint ($p = 1$) with $w^i = 1$, $i \in \mathcal{V}$ and all the decision variables have boxed inequality. On the other hand, Patrinos and Bemporad (2014) and Richter et al. (2011) propose numerical procedures to determine a bound on Lagrange multipliers for affine inequalities. As part of our contribution in this paper, we obtain an explicit closed-form upper-bound on the Lagrange multipliers of problem (1), which enables determining the size of the suitable penalty function weight for both exact and ϵ -exact penalty function methods.

2. Preliminaries

Let \mathbb{R} , $\mathbb{R}_{\geq 0}$, \mathbb{Z} , and $\mathbb{Z}_{>0}$ be, respectively, the set of real, non-negative real, integer, and positive integer numbers. For a given $i, j \in \mathbb{Z}$, $i < j$, we define $\mathbb{Z}_i^j = \{x \in \mathbb{Z} \mid i \leq x \leq j\}$. We denote the cardinality of a set \mathcal{A} by $|\mathcal{A}|$. For a matrix $\mathbf{A} = [a_{ij}] \in \mathbb{R}^{n \times m}$, we denote its transpose matrix by \mathbf{A}^\top , k th row by $[\mathbf{A}]_k$, k th column by $[\mathbf{A}]^k$, and its element wise max-norm with $\|\mathbf{A}\|_{\max}$. We let $\mathbf{1}_n$ (resp. $\mathbf{0}_n$) denote the vector of n ones (resp. n zeros), \mathbf{I}_n denote the $n \times n$ identity matrix. When clear from the context, we do not specify the matrix dimensions. For a vector $\mathbf{x} \in \mathbb{R}^n$ we denote the standard Euclidean and infinity norms by, respectively, $\|\mathbf{x}\| = \sqrt{\mathbf{x}^\top \mathbf{x}}$ and $\|\mathbf{x}\|_\infty = \max |x_i|_{i=1}^n$. Given a set of vectors, we use $[\{\mathbf{p}_i^i\}_{i \in \mathcal{M}}]$ to indicate the aggregate vector obtained from stacking the set of the vectors $\{\mathbf{p}_i^i\}_{i \in \mathcal{M}}$ whose indices belong to the ordered set $\mathcal{M} \subset \mathbb{Z}_{>0}$. In a network of N agents, to distinguish and emphasize that a variable is local to an agent $i \in \mathbb{Z}_1^N$, we use superscripts, e.g., $f^i(\mathbf{x}^i)$ is the local function of agent $i \in \mathbb{Z}_1^N$ evaluated at its own local value $\mathbf{x}^i \in \mathbb{R}^{n^i}$. The l th element of a vector $\mathbf{x}^i \in \mathbb{R}^{n^i}$ at agent $i \in \mathbb{Z}_1^N$ is denoted by x_l^i . Moreover, if $\mathbf{p}^i \in \mathbb{R}^{d^i}$ is a variable of agent $i \in \mathcal{V} = \{1, \dots, N\}$, the aggregated \mathbf{p}^i 's of the network is the vector $\mathbf{p} = [\{\mathbf{p}_i^i\}_{i \in \mathcal{V}}] = [\mathbf{p}^1, \dots, \mathbf{p}^N]^\top \in \mathbb{R}^d$. For a differentiable function $f : \mathbb{R}^d \rightarrow \mathbb{R}$, $\nabla f(\mathbf{x})$ represents its gradient. A differentiable function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ is convex (resp. α -strongly convex, $\alpha \in \mathbb{R}_{>0}$) over a convex set $C \subseteq \mathbb{R}^d$ if and only if $(\mathbf{z} - \mathbf{x})^\top (\nabla f(\mathbf{z}) - \nabla f(\mathbf{x})) \geq 0$ (resp. $\alpha \|\mathbf{z} - \mathbf{x}\|^2 \leq (\mathbf{z} - \mathbf{x})^\top (\nabla f(\mathbf{z}) - \nabla f(\mathbf{x}))$, or equivalently $\alpha \|\mathbf{z} - \mathbf{x}\| \leq \|\nabla f(\mathbf{z}) - \nabla f(\mathbf{x})\|$) for all $\mathbf{x}, \mathbf{z} \in C$. Moreover, it is strictly convex over a convex set $C \subseteq \mathbb{R}^d$ if and only if $(\mathbf{z} - \mathbf{x})^\top (\nabla f(\mathbf{z}) - \nabla f(\mathbf{x})) > 0$.

Next, we briefly review basic concepts from algebraic graph theory following Bullo, Cortés, and Martínez (2009). 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 $\mathbf{A} = [a_{ij}] \in \mathbb{R}^{N \times N}$ is a weighted adjacency matrix such that $a_{ij} > 0$ if $(i, j) \in \mathcal{E}$ and $a_{ij} = 0$, 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 whose weights satisfy $a_{ij} = a_{ji}$ for all $i, j \in \mathcal{V}$ is called a connected graph if there is a path from every node to every other node in the network.

The (out-)Laplacian matrix of a graph is $\mathbf{L} = \text{Diag}(\mathbf{A}\mathbf{1}_N) - \mathbf{A}$. Note that $\mathbf{L}\mathbf{1}_N = \mathbf{0}$. A graph is connected if and only if $\mathbf{1}_N^\top \mathbf{L} = \mathbf{0}$, and $\text{rank}(\mathbf{L}) = N - 1$. Therefore, for a connected graph zero is a simple eigenvalue of \mathbf{L} . For a connected graph, we denote the eigenvalues of \mathbf{L} by $\lambda_1, \dots, \lambda_N$, where $\lambda_1 = 0$ and $\lambda_i \leq \lambda_j$, for $i < j$.

3. Distributed continuous-time solvers

In this section, we present our distributed algorithm to first solve the constrained optimization problem (1) when there is no inequality constraint, i.e., $\mathcal{B}^i = \bar{\mathcal{B}}^i = \{\}$ for $i \in \mathcal{V}$. Then, we extend our results to solve the constrained optimization problem (1) with inequality constraints. Our standing assumptions are given below.

Assumption 3.1 (Problem Specifications). The cost function $f_l^i : \mathbb{R} \rightarrow \mathbb{R}$ of the subagent $l \in \mathbb{Z}_1^{n^i}$ of each agent $i \in \mathcal{V}$ is convex and differentiable. Moreover, $\nabla f^i : \mathbb{R}^{n^i} \rightarrow \mathbb{R}^{n^i}$ of each agent $i \in \mathcal{V}$ is locally Lipschitz. Also,

$$\mathbf{W} = [\mathbf{w}^1, \dots, \mathbf{w}^N] \in \mathbb{R}^{p \times m} \quad (2)$$

is full row rank and the feasible set

$$\mathbf{X}_{\text{fe}} = \{\mathbf{x} \in \mathbb{R}^m \mid (1\mathbf{b}), (1\mathbf{c}), (1\mathbf{d}) \text{ hold}\} \quad (3)$$

is non-empty for local inequalities (1c) and (1d). Lastly, the optimization problem (1) has a finite optimum $f^* = f(\mathbf{x}^*) = \sum_{i=1}^N f^i(\mathbf{x}^*)$. \square

Local Lipschitzness of ∇f^i , $i \in \mathcal{V}$, guarantees existence and uniqueness of the solution of our proposed algorithm (7), which is a differential equation.

To solve problem (1) subject to only the equality constraints, we consider the augmented cost function with a penalty term on violating the affine constraint, i.e.,

$$\mathbf{x}^* = \underset{\mathbf{x} \in \mathbb{R}^m}{\text{argmin}} \sum_{i=1}^N f^i(\mathbf{x}^i) + \frac{\rho}{2} \|\mathbf{W}\mathbf{x} - \mathbf{b}\|^2, \quad (4a)$$

$$[\mathbf{w}^1]_k \mathbf{x}^1 + \dots + [\mathbf{w}^N]_k \mathbf{x}^N = \mathbf{b}_k, \quad k \in \mathbb{Z}_1^p, \quad (4b)$$

where $\rho \in \mathbb{R}_{\geq 0}$ is the penalty parameter. This augmentation results in the so-called *augmented Lagrangian* formulation of iterative optimization algorithms. As stated in Boyd et al. (2010), augmented Lagrangian methods were developed in part to bring robustness to the dual ascent method, and in particular, to yield convergence without assumptions like strict convexity or finiteness of the cost function (see also Bertsekas & Tsitsiklis, 1997). As shown below, such positive effects are valid also for the continuous-time algorithms we study. Augmenting the cost with the penalty function as in (4a) however presents a challenge in design of distributed solutions as the total cost in (4a) is no longer separable. Nevertheless, we are able to address this challenge in our distributed solution.

Lemma 3.1 (KKT Conditions to Characterize Solution Set of (4) Boyd & Vandenberghe, 2004). Consider the constrained optimization problem (4). Let Assumption 3.1 hold and $f^i : \mathbb{R}^{n^i} \rightarrow \mathbb{R}$, $i \in \mathcal{V}$, be a differentiable and convex function on \mathbb{R}^{n^i} . For any $\rho \in \mathbb{R}_{\geq 0}$, a point $\mathbf{x}^* \in \mathbb{R}^m$ is a solution of (4) if and only if there exists a $\boldsymbol{\nu}^* \in \mathbb{R}^p$, such that, for $i \in \mathcal{V}$,

$$\nabla f^i(\mathbf{x}^*) + \mathbf{w}^{i\top} \boldsymbol{\nu}^* = \mathbf{0}, \quad (5a)$$

$$[\mathbf{w}^1]_k \mathbf{x}^1 + \dots + [\mathbf{w}^N]_k \mathbf{x}^N = \mathbf{b}_k, \quad k \in \mathbb{Z}_1^p. \quad (5b)$$

Moreover, $\boldsymbol{\nu}^*$ corresponding to every \mathbf{x}^* is unique and finite. If the local cost functions are strongly convex, then for any $\rho \in \mathbb{R}_{\geq 0}$ the KKT equation (5) has a unique solution $(\boldsymbol{\nu}^*, \mathbf{x}^*)$, i.e., (4) has a unique solution. \square

Let $L(\mathbf{v}, \mathbf{x}) = f(\mathbf{x}) + \frac{\rho}{2} \|\mathbf{w}^1 \mathbf{x}^1 + \cdots + \mathbf{w}^N \mathbf{x}^N - \mathbf{b}\|^2 + \mathbf{v}^\top (\mathbf{w}^1 \mathbf{x}^1 + \cdots + \mathbf{w}^N \mathbf{x}^N - \mathbf{b})$ be the augmented Lagrangian of the optimization problem (4). Following Arrow et al. (1958), a central solver for the optimal resource allocation problem (4) is

$$\dot{v}_k = \frac{\partial L(\mathbf{v}, \mathbf{x})}{\partial v_k} = [\mathbf{w}^1]_k \mathbf{x}^1 + \cdots + [\mathbf{w}^N]_k \mathbf{x}^N - b_k, \quad (6a)$$

$$\dot{\mathbf{x}}^i = -\frac{\partial L(\mathbf{v}, \mathbf{x})}{\partial \mathbf{x}^i} = -\nabla f^i(\mathbf{x}^i) - \sum_{j=1}^p [\mathbf{w}^i]_j^\top v_j - \rho \mathbf{w}^{i\top} (\mathbf{w}^1 \mathbf{x}^1 + \cdots + \mathbf{w}^N \mathbf{x}^N - \mathbf{b}), \quad (6b)$$

where $k \in \mathbb{Z}_1^p$, and $i \in \mathcal{V}$. The algorithm studied in Arrow et al. (1958) is for un-augmented Lagrangian, i.e., $\rho = 0$, and the guaranteed convergence holds only for strictly convex cost function $f(\mathbf{x})$. However, we can show that the central solver (6) with $\rho > 0$ is guaranteed to converge for convex cost function $f(\mathbf{x})$, as well (the details are omitted for brevity). A numerical example demonstrating this positive role is presented in Moradian and Kia (2020, Appendix B).

The source of coupling in (4) is the set of the equality constraints (4b), which appear in the central solver (6), as well. To design our distributed algorithm, we adapt the structural constitution of (6), but aim to create the coupling terms $[\mathbf{w}^1]_k \mathbf{x}^1 + \cdots + [\mathbf{w}^N]_k \mathbf{x}^N - b_k$, $k \in \mathbb{Z}_1^p$, in a distributed manner. We note that for every equality constraint $k \in \mathbb{Z}_1^p$, the coupling is among the set of agents $\mathcal{C}_k = \{i \in \mathcal{V} \mid [\mathbf{w}^i]_k \neq \mathbf{0}\}$. To have an efficient communication and computation resource management, we seek an algorithm that handles every coupled equality constraint among only those agents that are involved. In this regards, for every equality constraint $k \in \mathbb{Z}_1^p$, we let $\mathcal{G}_k(\mathcal{V}_k, \mathcal{E}_k)$ be a connected undirected subgraph of \mathcal{G} that contains the set of agents \mathcal{C}_k (see Fig. 1 for an example). We assume that $\mathcal{V}_k \subset \mathcal{V}$ is a monotonically increasing ordered set. It is very likely that the agents coupled through an equality constraint are geographically close, and thus in the communication range of each other. Nevertheless, \mathcal{V}_k , $k \in \mathbb{Z}_1^p$, may contain agents $i \in \mathcal{V}$ that have $[\mathbf{w}^i]_k = \mathbf{0}$ but are needed to make \mathcal{G}_k connected (see Fig. 1 for an example). We let $N_k = |\mathcal{V}_k|$, $k \in \mathbb{Z}_1^p$. In our distributed solution for (4), we also seek an algorithm that allows each agent to use a local penalty parameter $\rho^i \in \mathbb{R}_{>0}$, so we can eliminate the need to coordinate among the agents to choose the penalty parameter ρ . In what follows, we define $\mathcal{T}^i = \{j \in \mathbb{Z}_1^p \mid i \in \mathcal{V}_j\}$, $i \in \mathcal{V}$, and $\{\bar{b}_k^l\}_{l \in \mathcal{V}_k}$ such that $\sum_{l \in \mathcal{V}_k} \bar{b}_k^l = b_k$, for $k \in \mathbb{Z}_1^p$ (possible options include $\bar{b}_k^l = b_k / |\mathcal{C}_k|$, $l \in \mathcal{C}_k$ while $\bar{b}_k^j = 0$, $j \in \mathcal{V} \setminus \mathcal{C}_k$, or $\bar{b}_k^j = b_k$ for a particular agent $j \in \mathcal{V}_k$ and $\bar{b}_k^l = 0$ for any $l \in \mathcal{V} \setminus \{j\}$).

With the right notation at hand, our proposed distributed algorithm to solve optimization problem (4) is

$$\dot{v}_k^l = \beta_k \sum_{j \in \mathcal{V}_k} a_{lj}(v_k^l - v_k^j), \quad (7a)$$

$$\dot{v}_k^l = ([\mathbf{w}^l]_k \mathbf{x}^l - \bar{b}_k^l) - \beta_k \sum_{j \in \mathcal{V}_k} a_{lj}(v_k^l - v_k^j) - y_k^l, \quad (7b)$$

$$\begin{aligned} \dot{\mathbf{x}}^i &= -(1 + \rho^i) \nabla f^i(\mathbf{x}^i) - \rho^i \sum_{k \in \mathcal{T}^i} [\mathbf{w}^i]_k^\top ([\mathbf{w}^i]_k \mathbf{x}^i - \bar{b}_k^i) \\ &+ \rho^i \sum_{k \in \mathcal{T}^i} ([\mathbf{w}^i]_k^\top y_k^i) - (1 + \rho^i) \sum_{k \in \mathcal{T}^i} ([\mathbf{w}^i]_k^\top v_k^i), \end{aligned} \quad (7c)$$

with $\beta_k \in \mathbb{R}_{>0}$ and $\rho^i \in \mathbb{R}_{\geq 0}$ for $i \in \mathcal{V}$, $k \in \mathbb{Z}_1^p$ and $l \in \mathcal{V}_k$. To comprehend the connection with the centralized dynamical solver (6), take summation of (7a) and (7b) over every connected \mathcal{G}_k , $k \in \mathbb{Z}_1^p$ to obtain

$$\sum_{l \in \mathcal{V}_k} \dot{y}_k^l = 0 \implies \sum_{l \in \mathcal{V}_k} y_k^l(t) = \sum_{l \in \mathcal{V}_k} y_k^l(0), \quad (8a)$$

$$\sum_{l \in \mathcal{V}_k} v_k^l = [\mathbf{w}^1]_k \mathbf{x}^1 + \cdots + [\mathbf{w}^N]_k \mathbf{x}^N - b_k, \quad (8b)$$

which shows that for any $k \in \mathbb{Z}_1^p$, the dynamics of the sum of v_k^l 's duplicates the Lagrange multiplier dynamics (6a) of the central Augmented Lagrangian method. Therefore, in a convergent (7), ultimately for each $k \in \mathbb{Z}_1^p$, all the v_k^l 's converge to the same value indicating that ultimately every agent obtains a local copy of (6a) for any $k \in \mathbb{Z}_1^p$. On the other hand, if we factor out $(1 + \rho^i)$ from the right hand side of (7c) and exclude the third component, which is a technical term added to induce agreement between the agents, (7c) mimics the dynamics (6b) of the central Augmented Lagrangian solver.

Remark 3.1 (Benefits of Cluster-Based Approach). First we note that regardless of the size of n^i , in algorithm (7) we associate at most one copy of the Lagrange multiplier generator dynamics, i.e., (7a) and (7b), to every agent $i \in \mathcal{V}$. Specifically, every agent $i \in \mathcal{V}$, maintains $|\mathcal{T}^i| \leq p$ number of (7a) and (7b) pair dynamics and consequently has to broadcast the same number of variables to the network. In comparison, if we use the algorithms in Alghunaim et al. (2018), Cherukuri and Cortés (2015, 2016), Ding and Jovanovic (2018), Kia (2017b), Xiao and Boyd (2006), Yi et al. (2016) and Zhang and Zavlanos (2018), when $n^i > 1$, for any $i \in \mathcal{V}$, we need to treat each component of the i as an agent and assign a copy of a dynamics that generates the dual variable to every subagent $l \in \mathbb{Z}_1^{n^i}$. This results in a storage, computation and communication cost of order $n^i \times p$ per agent $i \in \mathcal{V}$. See our numerical examples for a comparison. Next, notice that algorithm (7) can always be implemented by using $\mathcal{G}_k = \mathcal{G}$, $k \in \mathbb{Z}_1^p$, where $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is the connected interaction topology that all the agents form. However, the flexibility to use a smaller cyber-layer formed by only the cluster of agents that are coupled by an equality constraint reduces the communication and computational cost of implementing Algorithm (7). Moreover, in some problems, to our numerical example in Section 4, the coupling equation is between the neighboring agents. In such cases, subgraphs \mathcal{G}_k can be easily formed. Moreover, as one can expect and our numerical example also highlights, using a smaller subgraph \mathcal{G}_k can results in a faster convergence for (7a) and (7b) dynamics and as a result a faster convergence for algorithm (7). \square

The equilibrium points of algorithm (7) when every \mathcal{G}_k , $k \in \mathbb{Z}_1^p$ is a connected graph is given by

$$\begin{aligned} \mathcal{S}_e = \left\{ \{(\mathbf{v}_k)_{k=1}^p, \{\mathbf{y}_k\}_{k=1}^p, \{\mathbf{x}^i\}_{i=1}^N\} \in \prod_{k=1}^p \mathbb{R}^{N_k} \times \prod_{k=1}^p \mathbb{R}^{N_k} \times \right. \\ \left. \prod_{i=1}^N \mathbb{R}^{n^i} \mid \mathbf{v}_k = \theta_k \mathbf{1}_{N_k}, \theta_k \in \mathbb{R}, \nabla f^i(\mathbf{x}^i) + \sum_{j \in \mathcal{T}^i} [\mathbf{w}^i]_j^\top \theta_j = \mathbf{0}, \right. \\ \left. \sum_{j=1}^N [\mathbf{w}^j]_k \mathbf{x}^j = b_k + \sum_{j \in \mathcal{V}_k} y_k^j, y_k^l = [\mathbf{w}^l]_k \mathbf{x}^l - \bar{b}_k^l, \right. \\ \left. i \in \mathcal{V}, l \in \mathcal{V}_k, k \in \mathbb{Z}_1^p \right\}. \end{aligned} \quad (9)$$

Due to (8a), if algorithm (7) is initialized such that $\sum_{l \in \mathcal{V}_k} y_k^l(0) = 0$, we have $\sum_{l \in \mathcal{V}_k} y_k^l(t) = \sum_{l \in \mathcal{V}_k} y_k^l(0)$ for $t \in \mathbb{R}_{\geq 0}$. In that case, if algorithm (7) converges to an equilibrium point $(\{\bar{\mathbf{v}}_k\}_{k=1}^p, \{\bar{\mathbf{y}}_k\}_{k=1}^p, \{\bar{\mathbf{x}}^i\}_{i=1}^N) \in \mathcal{S}_e$, we have $(\{\bar{\mathbf{v}}_k\}_{k=1}^p, \{\bar{\mathbf{y}}_k\}_{k=1}^p, \{\bar{\mathbf{x}}^i\}_{i=1}^N) = (\{\{[\mathbf{w}^i]_k \mathbf{x}^i - \bar{b}_k^i\}_{l \in \mathcal{V}_k}\}_{k=1}^p, \{\mathbf{v}_k^* \mathbf{1}_{N_k}\}_{k=1}^p, \{\mathbf{x}^{i*}\}_{i=1}^N)$, where $(\{\mathbf{x}^{i*}\}_{i=1}^N, \{\mathbf{v}_k^*\}_{k=1}^p)$ satisfies the KKT equation (5). The following theorem shows that indeed under the stated initialization, the algorithm (7) converges to a minimizer of optimization problem (4).

Theorem 3.1 (Asymptotic Convergence of (7) Over Connected Graphs When the Local Costs are Convex). Let every \mathcal{G}_k , $k \in \mathbb{Z}_1^p$, be a connected graph and [Assumption 3.1](#) hold. For every $k \in \mathbb{Z}_1^p$, suppose $\{\bar{b}_k^l\}_{l \in \mathcal{V}_k} \subset \mathbb{R}$ is defined such that $\sum_{l \in \mathcal{V}_k} \bar{b}_k^l = b_k$. Then, for each $i \in \mathcal{V}$, $l \in \mathcal{V}_k$, starting from $\mathbf{x}^i(0) \in \mathbb{R}^{n^i}$ and $y_k^l(0), v_k^l(0) \in \mathbb{R}$ with $\sum_{l \in \mathcal{V}_k} y_k^l(0) = 0$, the algorithm (7) for any $\rho^i \in \mathbb{R}_{>0}$, makes $t \mapsto (\{\mathbf{v}_k(t)\}_{k=1}^p, \{\mathbf{x}^i(t)\}_{i=1}^N)$ converge asymptotically to $(\{\mathbf{v}_k^* \mathbf{1}_{N_k}\}_{k=1}^p, \{\mathbf{x}^{i*}\}_{i=1}^N)$, where $(\{\mathbf{v}_k^*\}_{k=1}^p, \{\mathbf{x}^{i*}\}_{i=1}^N)$ is a point satisfying the KKT conditions (5) of problem (4). \square

For brevity, the proof of [Theorem 3.1](#) is given in [Moradian and Kia \(2020\)](#). The initialization condition $\sum_{l \in \mathcal{V}_k} y_k^l(0) = 0$ of [Theorem 3.1](#) is trivially satisfied by every agent $l \in \mathcal{V}_k$, $k \in \mathbb{Z}_1^p$, using $y_k^l(0) = 0$. The asymptotic convergence guarantee for algorithm (7) in [Theorem 3.1](#) is established for local convex cost functions. For such cost functions, similar to the centralized algorithm (6), (7) fails to converge when $\rho^i = 0$ for all $i \in \mathcal{V}$. Next, we show that if the local costs are strongly convex and have Lipschitz gradients then the convergence is in fact exponentially fast for $\rho^i \in \mathbb{R}_{>0}$ $i \in \mathcal{V}$. Recall that for strongly convex local cost functions, the minimizer of (4) is unique.

Theorem 3.2 (Exponential Convergence of (7) Over Connected Graphs When the Local Costs are Strongly Convex and Have Lipschitz Gradients). Let every \mathcal{G}_k , $k \in \mathbb{Z}_1^p$ be connected and [Assumption 3.1](#) hold. Also, assume each cost function f_l^i , $l \in \mathbb{Z}_1^{n^i}$, $i \in \mathcal{V}$, is m_l^i -strongly convex and has M_l^i -Lipschitz gradient. Let $m = \max\{\{m_l^i\}_{i=1}^N\} \in \mathbb{R}_{>0}$ and $M = \max\{\{M_l^i\}_{i=1}^N\} \in \mathbb{R}_{>0}$. Then, starting from $\mathbf{x}^i(0) \in \mathbb{R}^{n^i}$ and $y_k^l(0), v_k^l(0) \in \mathbb{R}$ for each $i \in \mathcal{V}$, $l \in \mathcal{V}_k$, and given $\sum_{l \in \mathcal{V}_k} y_k^l(0) = 0$ and $\sum_{l \in \mathcal{V}_k} \bar{b}_k^l = b_k$ in (7), the algorithm (7) makes $t \mapsto (\{\mathbf{v}_k(t)\}_{k=1}^p, \{\mathbf{x}^i(t)\}_{i=1}^N)$ converge exponentially fast to $(\{\mathbf{v}_k^* \mathbf{1}_{N_k}\}_{k=1}^p, \{\mathbf{x}^{i*}\}_{i=1}^N)$ for any $\rho^i \in \mathbb{R}_{>0}$, where $(\{\mathbf{v}_k^*\}_{k=1}^p, \{\mathbf{x}^{i*}\}_{i=1}^N)$ is the unique solution of the KKT conditions (5) of problem (4). Moreover, when $\rho^i = 0$ for an $i \in \mathcal{V}$, the convergence to the unique solution of the KKT conditions (5) is asymptotic. \square

The proof of [Theorem 3.2](#) is given in [Moradian and Kia \(2020\)](#). From this proof, we can observe that [Theorem 3.2](#) holds also for the problems that the decision variable of any subagent l of any agent i is multi-dimensional.

Remark 3.2 (The Convergence of (7) Over Dynamically Changing Connected Graphs). The proof of [Theorem 3.2](#) relies on a Lyapunov function that is independent of the systems parameters, and its derivative for $\rho^i \in \mathbb{R}_{>0}$, $i \in \mathcal{V}$, is negative definite with a quadratic upper bound. Hence, we can also show that the algorithm (7), when $\rho^i \in \mathbb{R}_{>0}$ for $i \in \mathcal{V}$, converges exponentially fast to a unique solution of the KKT conditions (5) of problem (4) over any time-varying topology \mathcal{G}_k , $k \in \mathbb{Z}_1^p$ that is connected at all times and its adjacency matrix is uniformly bounded and piece-wise constant.

3.1. Problem subject to both equality and inequality constraints

To address inequality constraints, we use a penalty function method to eliminate the local inequality constraints (1c) and (1d). That is, we seek solving

$$\mathbf{x}_p^* = \arg \min_{\mathbf{x} \in \mathbb{R}^m} \sum_{i=1}^N f_p^i(\mathbf{x}^i), \quad \text{subject to} \quad (10a)$$

$$[\mathbf{w}^1]_j \mathbf{x}^1 + \cdots + [\mathbf{w}^N]_j \mathbf{x}^N = b_j, \quad j \in \mathbb{Z}_1^p, \quad (10b)$$

with

$$f_p^i(\mathbf{x}^i) = f^i(\mathbf{x}^i) + \gamma \left(\sum_{l \in \underline{\mathcal{B}}^i} p_\epsilon(\underline{x}_l^i - x_l^i) + \sum_{l \in \bar{\mathcal{B}}^i} p_\epsilon(x_l^i - \bar{x}_l^i) \right), \quad (11)$$

$i \in \mathcal{V}$, where $\gamma \in \mathbb{R}_{>0}$ is the weight of the smooth penalty function $p_\epsilon = \begin{cases} 0, & y \leq 0, \\ \frac{1}{2\epsilon}y^2, & 0 \leq y \leq \epsilon, \\ (y - \frac{1}{2}\epsilon), & y \geq \epsilon, \end{cases}$ for some $\epsilon \in \mathbb{R}_{>0}$. This approach allows us to use algorithm (7) to solve the optimization (1) by using $f_p^i(\mathbf{x}^i)$ in place of $f^i(\mathbf{x}^i)$ in (7c). We note that $f_p^i(\mathbf{x}^i)$ is convex and differentiable if $f^i(\mathbf{x}^i)$ is a convex function in \mathbb{R}^{n^i} . Following this penalty method approach, when the global cost function of (1) is evaluated at the limit point of algorithm (7), it is in ϵ -order neighborhood of the global optimal value of the optimization problem (1) (see [Proposition 3.1](#)). In what follows, we investigate when the penalty function weight γ has a finite value and give a well-defined admissible range for it.

Given [Assumption 3.1](#), the Slater condition ([Boyd & Vandenberghe, 2004](#)) is satisfied. Thus, the KKT conditions below give a set of necessary and sufficient conditions that characterize the solution set of the convex optimization problem (1).

Lemma 3.2 (Solution Set of (1) [Boyd & Vandenberghe, 2004](#)). Consider the constrained optimization problem (1) under [Assumption 3.1](#). A point $\mathbf{x}^* \in \mathbb{R}^m$ is a solution of (1) if and only if there exists $\nu^* \in \mathbb{R}^p$ and $\{\underline{\mu}_l^{i*}\}_{l \in \underline{\mathcal{B}}^i} \subset \mathbb{R}_{\geq 0}$, $\{\bar{\mu}_l^{i*}\}_{l \in \bar{\mathcal{B}}^i} \subset \mathbb{R}_{\geq 0}$, $i \in \mathcal{V}$, such that

$$\nabla f^i(\mathbf{x}^{i*}) + \mathbf{w}^{i\top} \nu^* - \underline{\mu}^{i*} + \bar{\mu}^{i*} = \mathbf{0}, \quad (12a)$$

$$\mathbf{Wx}^* - \mathbf{b} = \mathbf{0}, \quad (12b)$$

$$\underline{\mu}_l^{i*}(\underline{x}_l^i - x_l^{i*}) = 0, \quad \underline{x}_l^i - x_l^{i*} \leq 0, \quad \underline{\mu}_l^{i*} \geq 0, \quad l \in \underline{\mathcal{B}}^i, \quad (12c)$$

$$\bar{\mu}_l^{i*}(\bar{x}_l^i - x_l^{i*}) = 0, \quad x_l^{i*} - \bar{x}_l^i \leq 0, \quad \bar{\mu}_l^{i*} \geq 0, \quad l \in \bar{\mathcal{B}}^i, \quad (12d)$$

where $\underline{\mu}^{i*} = [\underline{\mu}_1^{i*}, \dots, \underline{\mu}_{n^i}^{i*}]^\top$ with $\underline{\mu}_l^{i*} = 0$ for $l \in \mathbb{Z}_1^{n^i} \setminus \underline{\mathcal{B}}^i$ and $\bar{\mu}^{i*} = [\bar{\mu}_1^{i*}, \dots, \bar{\mu}_{n^i}^{i*}]^\top$ with $\bar{\mu}_l^{i*} = 0$ for $l \in \mathbb{Z}_1^{n^i} \setminus \bar{\mathcal{B}}^i$. If the local cost functions are strongly convex, then the optimization problem (1) has a unique solution. \square

Let X_{fe}^ϵ be the ϵ -feasible set of optimization problem (1),

$$X_{fe}^\epsilon = \left\{ \mathbf{x} \in \mathbb{R}^m \mid \mathbf{Wx} = \mathbf{b}, \quad \underline{x}_l^i - x_l^{i*} \leq \epsilon, \quad l \in \underline{\mathcal{B}}^i \right. \\ \left. x_j^i - \bar{x}_j^i \leq \epsilon, \quad j \in \bar{\mathcal{B}}^i, \quad i \in \mathcal{V} \right\}. \quad (13)$$

The result below states that for some admissible values of γ , the minimizer of problem (10) belongs to ϵ -feasible set X_{fe}^ϵ and optimal value of optimization problem (1) is in ϵ order neighborhood of the optimal value of the original optimization problem (1).

Proposition 3.1 (Relationship Between the Solution of (1) and (10) [Pinar & Zenios, 1994](#)). Let $(\mathbf{x}^*, \nu^*, \{\underline{\mu}_l^{i*}\}_{l \in \underline{\mathcal{B}}^i}, \{\bar{\mu}_l^{i*}\}_{l \in \bar{\mathcal{B}}^i})$ be any solution of the KKT Eqs. (5). Let \mathbf{x}_p^* be a minimizer of optimization problem (10) for some $\gamma, \epsilon \in \mathbb{R}_{>0}$. If $\gamma = \frac{1-N}{1-\sqrt{N}}\gamma^*$, where $\gamma^* > \max\{\max\{\underline{\mu}_l^{i*}\}_{l \in \underline{\mathcal{B}}^i}, \max\{\bar{\mu}_l^{i*}\}_{l \in \bar{\mathcal{B}}^i}\}_{i=1}^N$, then

$$\mathbf{x}_p^* \in X_{fe}^\epsilon, \quad 0 \leq f^* - f(\mathbf{x}_p^*) \leq \epsilon \gamma N, \quad (14)$$

where $f^* = f(\mathbf{x}^*)$ is the optimal value of (1). \square

Remark that by slight tightening of the inequality constraints according to $\underline{x}_l^i \leq \bar{x}_l^i - \epsilon$ and $\underline{x}_l^i + \epsilon \leq x_l^i$ and using these modified inequalities in the penalty function, we can guarantee that $\mathbf{x}_p^* \in X_{fe}$. But this may result in slight increase in the optimally gap in (14). Next, we note that if $\epsilon \rightarrow 0$, we have $p_\epsilon(y) \rightarrow p(y) = \max\{0, y\}$, where $p(y)$ is the well-known non-smooth penalty function ([Bertsekas, 1975](#)) with exact equivalency guarantees when $\gamma > \gamma^*$ in [Proposition 3.1](#).

Considering [Proposition 3.1](#), a practical and numerically well-posed solution via the penalty optimization method (10) is

achieved when the Lagrange multipliers are bounded. Thus, in what follows we seek for μ_{bound} in

$$\max \left\{ \max \{\underline{\mu}_l^{i*}\}_{l \in \underline{\mathcal{B}}^i}, \max \{\bar{\mu}_l^{i*}\}_{l \in \bar{\mathcal{B}}^i} \right\}_{i=1}^N \leq \mu_{\text{bound}}, \quad (15)$$

with the objective of choosing a penalty function weight γ that satisfies the condition set by [Proposition 3.1](#) by setting $\gamma \geq \frac{1-N}{1-\sqrt{N}} \mu_{\text{bound}}$.

For any solution of the KKT conditions (5), we let $\underline{\mathcal{A}}^i \subset \underline{\mathcal{B}}^i$ and $\bar{\mathcal{A}}^i \subset \bar{\mathcal{B}}^i$ respectively be the set of indices of the active lower bound and the active upper bound inequality constraints of agent $i \in \mathcal{V}$. We note that $\underline{\mathcal{A}}^i \cap \bar{\mathcal{A}}^i = \{\}$. Because for inactive inequalities $\bar{\mu}_l^{i*} = 0$ (resp. $\underline{\mu}_l^{i*} = 0$) for $l \in \bar{\mathcal{B}}^i \setminus \bar{\mathcal{A}}^i$ and $i \in \mathcal{V}$ (resp. $l \in \underline{\mathcal{B}}^i \setminus \underline{\mathcal{A}}^i$) ([Bertsekas, 1999](#)), we obtain

$$\begin{aligned} \max \left\{ \max \{\underline{\mu}_l^{i*}\}_{l \in \underline{\mathcal{B}}^i}, \max \{\bar{\mu}_l^{i*}\}_{l \in \bar{\mathcal{B}}^i} \right\}_{i=1}^N = \\ \max \left\{ \max \{\underline{\mu}_l^{i*}\}_{l \in \underline{\mathcal{A}}^i}, \max \{\bar{\mu}_l^{i*}\}_{l \in \bar{\mathcal{A}}^i} \right\}_{i=1}^N. \end{aligned} \quad (16)$$

Therefore, to find μ_{bound} , it suffices to find an upper bound on $\max \left\{ \max \{\underline{\mu}_l^{i*}\}_{l \in \underline{\mathcal{A}}^i}, \max \{\bar{\mu}_l^{i*}\}_{l \in \bar{\mathcal{A}}^i} \right\}_{i=1}^N$.

As known, the set of the Lagrange multipliers of an optimization problem of form (1) is nonempty and bounded if and only if the Mangasarian–Fromovitz constraint qualification (MFCQ) holds ([Mangasarian & Fromovitz, 1967](#)). It is straight-forward to show that the MFCQ condition is satisfied for a resource allocation problem of form (1) with one equality constraint (i.e., $p = 1$) and upper and lower bounded decision variables (i.e., $\underline{\mathcal{B}}^i = \bar{\mathcal{B}}^i = \mathbb{Z}_1^{n_i}$). For such a problem the following result specifies a μ_{bound} that satisfies (15).

Proposition 3.2 (μ_{bound} for the Resource Allocation Problem with One Equality Constraint and Bounded Decision Variables). Consider problem (1) under [Assumption 3.1](#) when $p = 1$, $w_l^i > 0$ for $l \in \{1, \dots, n_i^i\}$ and $\underline{\mathcal{B}}^i = \bar{\mathcal{B}}^i = \mathbb{Z}_1^{n_i^i}$, $i \in \mathcal{V}$. Let $(\mathbf{x}^*, v^*, \{\underline{\mu}_l^{i*}\}_{l \in \underline{\mathcal{B}}^i}, \{\bar{\mu}_l^{i*}\}_{l \in \bar{\mathcal{B}}^i})$ be an arbitrary solution of the KKT conditions (5) for this problem. Then, μ_{bound} in (15) satisfies

$$\mu_{\text{bound}} \leq (1 + \frac{\bar{w}}{w}) \max \left\{ \max_{\mathbf{x}^i \in X_{\text{ineq}}^i} \|\nabla f^i(\mathbf{x}^i)\|_{\infty} \right\}_{i=1}^N, \quad (17)$$

where $X_{\text{ineq}}^i = \{\mathbf{x}^i \in \mathbb{R}^{n_i^i} \mid \underline{x}_l^i \leq x_l^i \leq \bar{x}_l^i, l \in \mathbb{Z}_1^{n_i^i}\}$, $w = \min \{\{w_l^i\}_{l=1}^{n_i^i}\}_{i=1}^N$ and $\bar{w} = \max \{\{w_l^i\}_{l=1}^{n_i^i}\}_{i=1}^N$.

Proof. For any given $(\mathbf{x}^*, v^*, \{\underline{\mu}_l^{i*}\}_{l \in \underline{\mathcal{B}}^i}, \{\bar{\mu}_l^{i*}\}_{l \in \bar{\mathcal{B}}^i})$, we note that the KKT conditions (5) can be written as

$$\nabla f_l^i(x_l^{i*}) + w_l^i v^* = 0, \quad l \in \mathbb{Z}_1^{n_i^i} \setminus (\bar{\mathcal{A}}^i \cup \underline{\mathcal{A}}^i), \quad (18a)$$

$$\nabla f_l^i(x_l^{i*}) + w_l^i v^* + \bar{\mu}_l^{i*} = 0, \quad l \in \bar{\mathcal{A}}^i, \quad (18b)$$

$$\nabla f_l^i(x_l^{i*}) + w_l^i v^* - \underline{\mu}_l^{i*} = 0, \quad l \in \underline{\mathcal{A}}^i. \quad (18c)$$

Since $\{w_l^i\}_{l=1}^{n_i^i} \subset \mathbb{R}_{>0}$, it follows from [Assumption 3.1](#), which states that the feasible set is non-empty for strict local inequalities, that the upper bounds (similarly the lower bounds) for all decision variable cannot be active simultaneously. Therefore, for any given minimizer, we have either (a) at least for one subagent $k \in \mathbb{Z}_1^{n_i^i}$ in an agent $i \in \mathcal{V}$ we have $\underline{x}_k^i < x_k^{i*} < \bar{x}_k^i$ or (b) some of the decision variables are equal to their upper bound and the remaining others are equal to their lower bound. If case (a) holds, it follows from (18a) that $v^* = \frac{-\nabla f_k^i(x_k^{i*})}{w_k^i}$, which means that we

have the guarantees that $|v^*| \leq \frac{\max \{\|\nabla f^i(\mathbf{x}^i)\|_{\infty}\}_{i=1}^N}{w}$. On the other hand, if (b) holds, then there exists at least an agent $k \in \mathcal{V}$ with $\bar{\mathcal{A}}^k \neq \{\}$ and an agent $j \in \mathcal{V}$ with $\underline{\mathcal{A}}^j \neq \{\}$ ($k = j$ is possible). Therefore, for $l \in \bar{\mathcal{A}}^k$ it follows from (18b) that $v^* = \frac{1}{w_k^k} (-\nabla f_l^k(x_l^{k*}) - \bar{\mu}_l^{k*})$, and for $l \in \underline{\mathcal{A}}^j$ it follows from (18c) that $v^* = \frac{1}{w_j^j} (-\nabla f_l^j(x_l^{j*}) + \bar{\mu}_l^{j*})$. Consequently, because $\bar{\mu}_l^{k*} \geq 0$ and $\bar{\mu}_l^{j*} \geq 0$, we conclude that $-\frac{1}{w_k^k} \nabla f_l^k(x_l^{k*}) \leq v^* \leq -\frac{1}{w_j^j} \nabla f_l^j(x_l^{j*})$, which leads to $|v^*| \leq \max \{|\frac{\nabla f_l^k(x_l^{k*})}{w_k^k}|, |\frac{\nabla f_l^j(x_l^{j*})}{w_j^j}|\} \leq \frac{\max \{\|\nabla f^i(\mathbf{x}^i)\|_{\infty}\}_{i=1}^N}{w}$. Therefore, we conclude that for any given $(\mathbf{x}^*, v^*, \{\underline{\mu}_l^{i*}\}_{l \in \underline{\mathcal{B}}^i}, \{\bar{\mu}_l^{i*}\}_{l \in \bar{\mathcal{B}}^i})$, we have $|v^*| \leq \frac{\max \{\|\nabla f^i(\mathbf{x}^i)\|_{\infty}\}_{i=1}^N}{w} \leq \frac{\max \left\{ \max_{\mathbf{x}^i \in X_{\text{ineq}}^i} \|\nabla f^i(\mathbf{x}^i)\|_{\infty} \right\}_{i=1}^N}{w}$. Consequently, it follows from (18b) that $\bar{\mu}_l^{i*} \leq |\nabla f_l^i(x_l^{i*})| + |w_l^i v^*| \leq \|\nabla f_l^i(x_l^{i*})\|_{\infty} + \bar{w} |v^*|$, and from (18c) that $\underline{\mu}_l^{i*} \leq |\nabla f_l^i(x_l^{i*})| + |w_l^i v^*| \leq \|\nabla f_l^i(x_l^{i*})\|_{\infty} + \bar{w} |v^*|$. Therefore, given (16), we have the guarantees that (17) holds. \square

To compute the upper-bound in (17) in a distributed way, agents can run a set of max-consensus algorithms. Evaluating the MFCQ condition generally is challenging for other classes of optimization problems. A common sufficient condition for the MFCQ is the linear independence constraint qualification (LICQ), which also guarantees the uniqueness of the Lagrange multipliers for any solution of the optimization problem (1) ([Wachsmuth, 2013](#)) (see [Wang & Elia, 2011](#) and [Srivastava & Cortes, 2018](#) for examples of the optimization solvers that are developed under the assumption that the LICQ holds).

We evaluate μ_{bound} for the general form of (1) by assuming LICQ condition in our extended version ([Moradian & Kia, 2020](#)).

4. Numerical examples

We consider a simple distributed self-localizing deployment problem concerned with optimal deployment of 3 sensors labeled S^i , $i \in \{1, 3, 5\}$ on a line to monitor a set of events that are horizontally located at $\mathbf{P} = \{[p_i]_{i=1}^{10}\} = [12, 11, 9, 3, 2, -1, -2, -8, -11, -13]$ for $t \in [0, 100]$, and $\mathbf{P} = \{[p_i]_{i=1}^{10}\} = [24, 22, 17, 15, 13, 8, 7, 3, -2, -4]$ for $t \in [100, 200]$, see [Fig. 2](#). Agent 1 is monitoring $\{p_i\}_{i=1}^3$, agent 3 is monitoring $\{p_i\}_{i=4}^7$, and agent 5 is monitoring $\{p_i\}_{i=8}^{10}$. Sensors should find their positions cooperatively to keep their position in the communication range of each other as well as stay close to the targets to improve the detection accuracy. Due to limited communication range, two relay nodes R^i , $i \in \{2, 4\}$, as shown in [Fig. 2](#) are used to guarantee the connectivity of the sensors during the operation. The problem is formulated by

$$\begin{aligned} \mathbf{x}^* = \arg \min_{\mathbf{x} \in \mathbb{R}^5} \sum_{i=1}^5 f^i(\mathbf{x}^i), \quad \text{subject to} \\ x^j - x^{j+1} \leq 5, \quad j \in \{1, \dots, 4\}, \end{aligned} \quad (19)$$

where $f^i(\mathbf{x}^i) = \sum_{j \in E^i} \|x^i - p_j\|^2$ for $i \in \{1, 3, 5\}$ with $E^1 = \{1, \dots, 3\}$, $E^3 = \{4, \dots, 7\}$ and $E^5 = \{8, \dots, 10\}$ and $f^i(\mathbf{x}^i) = 0$ for $i \in \{2, 4\}$. Here, x^i with $i \in \{1, 3, 5\}$ (resp. $i \in \{2, 4\}$) is the horizontal position of sensor S^i (resp. relay node R^i). To transform problem (19) to the standard form described in (1) we introduce slack variables $x_2^i \in \mathbb{R}$ with $i \in \{1, \dots, 4\}$, to rewrite (19) as

$$\begin{aligned} \mathbf{x}^* = \arg \min_{\mathbf{x} \in \mathbb{R}^9} \sum_{i=1}^5 f^i(\mathbf{x}^i), \quad \text{subject to} \\ x_1^j - x_1^{j+1} + x_2^j = 5, \quad x_2^j \geq 0, \quad j \in \{1, \dots, 4\}, \end{aligned} \quad (20)$$

where $\mathbf{x}^i \in \mathbb{R}^2$ for $i \in \{1, 2, 3, 4\}$, $\mathbf{x}^5 \in \mathbb{R}$, and $f^i(\mathbf{x}^i) = f^i(x_1^i)$ for any $i \in \{1, \dots, 5\}$, i.e., $f^i(x_2^i) = 0$. We can run algorithm (7) by choosing the cyber layer equivalent to the physical connected topology between all the agent, i.e., $\mathcal{G}_k = \mathcal{G}$ for $k \in \{1, 2, 3, 4\}$, where \mathcal{G} is the line graph connecting all 5 agents. However, as stated earlier this configuration leads to extra computational and

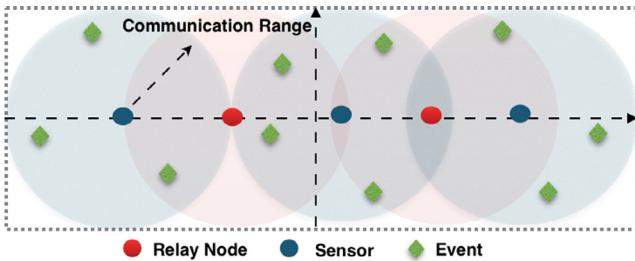


Fig. 2. Schematic representation of the events, sensors and relay nodes in the second example.

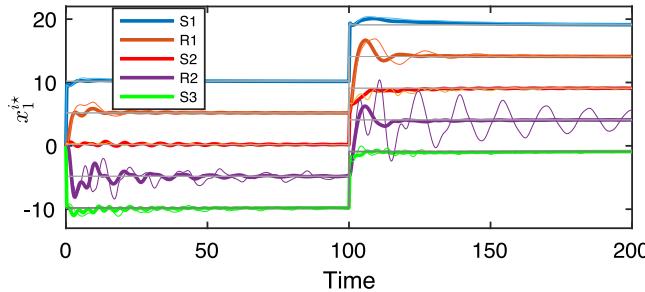


Fig. 3. Trajectories of $\{x_1^i\}_{i=1}^5$ generated by implementing distributed algorithm (7): The gray lines show the optimum positions of agents on the line obtained by using the Matlab's fmincon. The thick curved lines show the trajectories when algorithm (7) is implemented over cluster-based cyber-layers. the thin lines show the trajectories when algorithm (7) is implemented with $\mathcal{G}_k = \mathcal{G}$, $k = \{1, 2, 3, 4\}$.

communication efforts. Here, instead, we form 4 cyber-layers \mathcal{G}_k , $k \in \{1, 2, 3, 4\}$, where $\mathcal{V}_1 = \{1, 2\}$, $\mathcal{V}_2 = \{2, 3\}$, $\mathcal{V}_3 = \{3, 4\}$ and $\mathcal{V}_4 = \{4, 5\}$. We note that our proposed approach to form the cyber-layers in correspondence to the equality constraints leads to an efficient communication topology here. More specifically, to generate the dual dynamics, the agents $\{1, \dots, 5\}$, maintain and communicate variables of order $\{1, 2, 2, 2, 1\}$, respectively. Whereas, if we implement algorithms of Kia (2017b) and Zhang and Zavlanos (2018), the corresponding variables to generate the dual dynamics is of order $\{8, 8, 8, 8, 4\}$.

Fig. 3 shows the trajectory of the distributed optimization algorithm (7) (using $f_p^i(\mathbf{x}_1^i)$ as defined in (11) in place of $f^i(\mathbf{x}^i)$ in (7c)) for problem (20). As shown the location of the sensors remain in their communication range and converge to optimum values during execution of the algorithm (the optimal solution is shown by the gray lines, and is obtained by MATLAB's constraint optimization solver 'fmincon'). For the smooth penalty function (11) we use $\epsilon = 0.01$ and $\gamma = 200$, which satisfies the condition of Proposition 3.1. What is interesting to note in Fig. 3 is how the convergence of the algorithm is slowed down when we use $\mathcal{G}_k = \mathcal{G}$ for $k \in \mathbb{Z}_1^4$. This is expected, as in this case the coordination to generate the dual variables has to happen over a larger graph. A second simulation study, solving the optimal generator dispatch problem stated in Fig. 1, is available in Moradian and Kia (2020).

5. Conclusions

We proposed a novel cluster-based distributed augmented Lagrangian algorithm for a class of constrained convex optimization problems. In the design, we paid special attention to the efficient communication and computation resource management and required only the agents that are coupled through an equality constraint to form a communication topology to address that

coupling in a distributed manner. We showed that if the communication topology corresponding to each equality constraint is a connected graph, the proposed algorithm converges asymptotically when the local cost functions are convex, and exponentially when the local cost functions are strongly convex and have Lipschitz gradients. We invoked the ϵ -exact penalty function method to address the inequality constraints and obtained an explicit lower bound on the penalty function weight to guarantee convergence to ϵ -neighborhood of the global minimum value of the cost. Simulations demonstrated the performance of our proposed algorithm. As future work, we will study the event-triggered communication implementation of our algorithm.

References

Alghunaim, S. A., Yuan, K., & Sayed, A. H. (2018). Dual coupled diffusion for distributed optimization with affine constraints. In *IEEE Conf. on decision and control*. FL, USA.

Arrow, K. J., Hurwicz, L., & Uzawa, H. (1958). *Studies in linear and nonlinear programming*.

Bertsekas, D. P. (1975). Nondifferentiable optimization via approximation. *Mathematical Programming Study*, 3, 1–25.

Bertsekas, D. (1999). *Nonlinear programming*.

Bertsekas, D., & Tsitsiklis, J. (1997). *Parallel and distributed computation: Numerical methods*.

Boyd, S., Parikh, N., Chu, E., Peleato, B., & Eckstein, J. (2010). Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3, 1–122.

Boyd, S., & Vandenberghe, L. (2004). *Convex optimization*. Cambridge University Press.

Bullo, F., Cortés, J., & Martínez, S. (2009). *Applied mathematics series, Distributed control of robotic networks*. Princeton University Press.

Chen, J., & Lau, V. K. N. (2012). Convergence analysis of saddle point problems in time varying wireless systems – control theoretical approach. *IEEE Transactions on Signal Processing*, 60(1), 443–452.

Cherukuri, A., & Cortés, J. (2015). Distributed generator coordination for initialization and anytime optimization in economic dispatch. *IEEE Transactions on Control of Network Systems*, 2(3), 226–237.

Cherukuri, A., & Cortés, J. (2016). Initialization-free distributed coordination for economic dispatch under varying loads and generator commitment. *Automatica*, 74, 183–193.

Cherukuri, A., & Cortés, J. (2016). Initialization-free distributed coordination for economic dispatch under varying loads and generator commitment. *Automatica*, 74, 183–193.

Ding, D., Hu, B., Dhingra, N., & Jovanovic, M. (2018). An exponentially convergent primal-dual algorithm for nonsmooth composite minimization. In *IEEE Conf. on decision and control*. FL, USA.

Ding, D., & Jovanovic, M. (2018). A primal-dual Laplacian gradient flow dynamics for distributed resource allocation problems. In *American control conference*. WI, USA.

Domínguez-García, A. D., Cady, S. T., & Hadjicostis, C. N. (2012). Decentralized optimal dispatch of distributed energy resources. In *IEEE Conf. on decision and control* (pp. 3688–3693). Hawaii, USA.

Duchi, J., Agarwal, A., & Wainwright, M. (2012). Dual averaging for distributed optimization: Convergence analysis and network scaling. *IEEE Transactions on Automatic Control*, 57(3), 592–606.

Ferragut, A., & Paganini, F. (2014). Network resource allocation for users with multiple connections: fairness and stability. *IEEE/ACM Transactions on Networking*, 22(2), 349–362.

Haddad, W., & Chellaboina, V. (2008). *Nonlinear dynamical systems and control*. Princeton University Press.

Jakovetic, D., Moura, J., & Xavier, J. (2015). Linear convergence rate of a class of distributed augmented Lagrangian algorithms. *IEEE Transactions on Automatic Control*, 60(4), 922–936.

Kar, S., & Hug, G. (2012). Distributed robust economic dispatch in power systems: A consensus + innovations approach. In *Power & energy society general meeting* (pp. 1–8). San Diego, CA.

Kia, S. S. (2016). Distributed optimal resource allocation over networked systems and use of an epsilon-exact penalty function. In *IFAC symposium on large scale complex systems*. CA, USA.

Kia, S. S. (2017a). An augmented Lagrangian distributed algorithm for an in-network optimal resource allocation problem. In *American control conference*. WA, USA.

Kia, S. S. (2017b). Distributed optimal in-network resource allocation algorithm design via a control theoretic approach. *Systems & Control Letters*, 107, 49–57.

Kia, S. S., Cortés, J., & Martínez, S. (2014). Distributed convex optimization via continuous-time coordination algorithms with discrete-time communication. *Automatica*, 55, 254–264.

Madan, R., & Lall, S. (2006). Distributed algorithms for maximum lifetime routing in wireless sensor networks. *IEEE Transactions on Wireless Communication*, 5(8), 2185–2193.

Mangasarian, O. (1985). Computable numerical bounds for LAGRANGE multipliers of stationary points of non-convex differentiable non-linear programs. *Operations Research Letters*, 4(2), 1757–1780.

Mangasarian, O. L., & Fromovitz, S. (1967). The Fritz John necessary optimality conditions in the presence of equality and inequality constraints. *Operations Research Letters*, 17, 37–47.

Moradian, H., & Kia, S. S. (2020). Cluster-based distributed augmented Lagrangian algorithm for a class of constrained convex optimization problems. Available on [arXiv:1908.06634](https://arxiv.org/abs/1908.06634).

Patrinos, P., & Bemporad, A. (2014). An accelerated dual gradient-projection algorithm for embedded linear model predictive control. *IEEE Transactions on Automatic Control*.

Pinar, M. Ç., & Zenios, S. A. (1994). On smoothing exact penalty functions for convex constrained optimization. *IEEE Transactions on Communications*, 4(3), 1136–1144.

Richter, S., Morari, M., & Jones, C. (2011). Towards computational complexity certification for constrained MPC based on Lagrange relaxation and the fast gradient method. In *IEEE Conf. on decision and control* (pp. 5223–5229). Orlando, Florida, USA.

Rostami, R., Costantini, G., & Görges, D. (2017). ADMM-based distributed model predictive control: Primal and dual approaches. In *IEEE Conf. on decision and control*. Melbourne, Australia.

Srivastava, P., & Cortes, J. (2018). Distributed algorithm via continuously differentiable exact penalty method for network optimization. In *IEEE Conf. on decision and control*. FL, USA.

Vaquero, M., & Cortes, J. (2018). Distributed augmentation-regularization for robust online convex optimization. *IFAC-PapersOnLine*, 51(23), 230–235.

Varagnolo, D., Zanella, F., Cenedese, A., Pillonetto, G., & Schenato, L. (2015). Newton-Raphson consensus for distributed convex optimization. *IEEE Transactions on Automatic Control*, 61(4), 994–1009.

Wachsmuth, G. (2013). On LICQ and the uniqueness of Lagrange multipliers. *Operations Research Letters*, 41(1), 78–80.

Wang, J., & Elia, N. (2011). A control perspective for centralized and distributed convex optimization. In *IEEE Conf. on Decision and Control*. FL, USA.

Wei, W., Wang, J., Li, N., & Mei, S. (2017). Optimal power flow of radial networks and its variations: A sequential convex optimization approach. *IEEE Transactions on Smart Grid*, 8(6), 2974–2987.

Wood, A. J., Wollenberg, F., & Sheble, G. B. (2013). *Power generation, operation and control* (3rd ed.). New York: John Wiley.

Xiao, L., & Boyd, S. (2006). Optimal scaling of a gradient method for distributed resource allocation. *Journal of Optimization Theory and Applications*, 129(3), 469–488.

Xiao, L., Johansson, M., & Boyd, S. P. (2004). Simultaneous routing and resource allocation via dual decomposition. *IEEE Transactions on Communications*, 52(7), 1136–1144.

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

Zhang, Z., & Chow, M. (2012). Convergence analysis of the incremental cost consensus algorithm under different communication network topologies in a smart grid. *IEEE Transactions on Power Systems*, 27(4), 1761–1768.

Zhang, Y., & Zavlanos, M. M. (2018). A consensus-based distributed augmented Lagrangian method. In *IEEE Conf. on decision and control*. CA, USA.

Zholbarysov, M., Fooladivanda, D., & Domínguez-García, A. D. (2019). Resilient distributed optimal generation dispatch for lossy ac microgrids. *Systems & Control Letters*, 123, 47–54.



Hossein Moradian is a Ph.D. candidate in the Department of Mechanical and Aerospace Engineering, University of California, Irvine (UCI). He obtained his M.Sc. and B.Sc. in Mechanical Engineering from the Sharif University of Technology, Iran, in 2011 and 2014, respectively. His research interests include distributed optimization, statistical learning, nonlinear control theory and robotics.



Solmaz S. Kia is an Associate Professor in the Department of Mechanical and Aerospace Engineering, University of California, Irvine (UCI). She has a joint appointment with the Computer Science Department of UCI. She obtained her Ph.D. degree in Mechanical and Aerospace Engineering from UCI, in 2009, and her M.Sc. and B.Sc. in Aerospace Engineering from the Sharif University of Technology, Iran, in 2004 and 2001, respectively. She was a senior research engineer at SySense Inc., El Segundo, CA from Jun. 2009–Sep. 2010. She held postdoctoral positions in the Department of Mechanical and Aerospace Engineering at the UC San Diego and UCI. She was the recipient of UC President's Postdoctoral Fellowship from 2012–2014. She is also a recipient of 2017 NSF CAREER award. Dr. Kia is a senior member of IEEE, and serves as associate editor of IEEE Sensors Letters. Dr. Kia's main research interests, in a broad sense, include nonlinear control theory, distributed optimization/coordination/estimation, and probabilistic robotics.