Automatica 185 (2026) 112774

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Achieving violation-free distributed optimization under coupling constraints[☆]

Changxin Liu [a,*], Xiao Tan [b], Xuyang Wu [c], Dimos V. Dimarogonas [b], Karl H. Johansson [b]

[a] *Key Laboratory of Smart Manufacturing in Energy Chemical Process, Ministry of Education, East China University of Science and Technology, Shanghai 200237, China*
[b] *School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, and Digital Futures, SE-100 44, Stockholm, Sweden*
[c] *Shenzhen Key Laboratory of Control Theory and Intelligent Systems, Southern University of Science and Technology, Shenzhen 518055, China*

ARTICLE INFO

Article history:
Received 24 February 2024
Received in revised form 15 December 2024
Accepted 13 November 2025
Available online 18 December 2025

Keywords:
Distributed optimization
Constrained optimization
All-time constraint satisfaction
Safe distributed control

ABSTRACT

Constraint satisfaction is a critical component in a wide range of engineering applications, including safe multi-agent control and economic dispatch in power systems. This study explores violation-free distributed optimization techniques for problems characterized by separable objective functions and coupling constraints. First, we incorporate auxiliary decision variables together with a network-dependent linear mapping to each coupling constraint. For the reformulated problem, we show that the projection of its feasible set onto the space of primal variables is identical to that of the original problem, which is the key to achieving all-time constraint satisfaction. Upon treating the reformulated problem as a min-min optimization problem with respect to auxiliary and primal variables, we demonstrate that the gradients in the outer minimization problem have a locally computable closed-form. Then, two violation-free distributed optimization algorithms are developed and their convergence under reasonable assumptions is analyzed. Finally, the proposed algorithm is applied to implement a control barrier function based controller in a distributed manner, and the results verify its effectiveness.

1. Introduction

Distributed optimization problems over networks, such as economic dispatch in power networks and coordination in multi-agent systems, typically involve decision variables that are subject to coupling constraints. These constraints typically represent the shared resources of the parties involved and cannot be violated because of physical limitations. Due to its paramount importance, the design of distributed optimization algorithms for constraint-coupled problems has received increasing attention lately (see, for example, Camisa, Farina, Notarnicola, & Notarstefano, 2021; Li, Feng, & Xie, 2020; Notarnicola & Notarstefano, 2019; Tan & Dimarogonas, 2021; Wu, Magnússon, & Johansson, 2023).

An essential requirement in designing distributed optimization algorithms for such problems is all-time satisfaction of the coupling constraints. This is because the constraint satisfaction is critical to ensuring that the system operates safely and effectively (Ames, Xu, Grizzle, & Tabuada, 2016). Furthermore, all-time satisfaction guarantees that the iterative algorithm ends up with a feasible and safe-to-implement solution whenever it is stopped. Nevertheless, existing works typically focus on the convergence speed of the algorithm and overlook the requirement on constraint satisfaction during execution. Indeed, most existing results only have asymptotic feasibility guarantee, that is, the constraint violation vanishes asymptotically (Camisa et al., 2021; Li et al., 2020; Notarnicola & Notarstefano, 2019). Projecting such solutions onto the feasible set in presence of coupling constraints is usually not permissible in a distributed scenario. To this end, this work aims at developing violation-free distributed algorithms for optimization problems with separable objective functions and coupling constraints.

1.1. Related works

The most well-known methodology for solving constraint-coupled distributed optimization problems is dual decomposition (Boyd, Parikh, Chu, Peleato, Eckstein, et al., 2011). For

---

[☆] This work was supported in part by the National Natural Science Foundation of China under Grant 62573196, in part by the State Key Laboratory of Industrial Control Technology, China, under Grant ICT2025A02, and in part by the Swedish Research Council, the Knut and Alice Wallenberg Foundation, and the Swedish Foundation for Strategic Research. The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Kok Lay Teo under the direction of Editor Florian Dorfler.

[*] Corresponding author.

E-mail addresses: changxinl@ecust.edu.cn (C. Liu), xiaotan@kth.se (X. Tan), wuxy6@sustech.edu.cn (X. Wu), dimos@kth.se (D.V. Dimarogonas), kallej@kth.se (K.H. Johansson).

https://doi.org/10.1016/j.automatica.2025.112774

instance, the authors in Falsone, Margellos, Garatti, and Prandini (2017), Liu, Li, and Shi (2020) and Nedić and Ozdaglar (2009) considered the Lagrangian dual of the original problem and developed algorithms for distributed optimization problems with general separable constraints. Based on a similar saddle-point formulation, the authors in Mateos-Núnez and Cortés (2016) proposed distributed primal–dual algorithms, where the primal variables are updated via one-step gradient descent rather than solving subproblems. For the case with coupling equality constraints, Li, Xie, and Hong (2019) developed a continuous-time distributed optimization algorithm. For the generalized Nash equilibrium seeking problem under coupling constraints, Meng and Li (2023) designed a distributed primal–dual algorithm. Notably, the authors in Notarnicola and Notarstefano (2019) considered a relaxed version of the constraint-coupled optimization problem, based on which a distributed algorithm was designed to ensure the last iterate convergence of variables without averaging steps. Utilizing the same methodology, the authors in Wang, Papachristodoulou and Margellos (2023) developed an improved optimization algorithm for strongly convex problems. For sparsely coupled equality constraints, the authors in Alghunaim, Yuan, and Sayed (2019) introduced a structure-aware Lagrangian dual formulation, based on which the overall algorithm can be made more efficient. However, the primal iterates by all the above algorithms only satisfy the coupling constraints asymptotically, that is, the constraint violation asymptotically vanishes. Therefore, they may not generate feasible solutions within finite time, and are not implementable to safety-critical systems.

Another strategy for designing constraint-coupled distributed optimization is primal decomposition (Tan & Dimarogonas, 2021; Wu, Magnússon, & Johansson, 2021; Wu et al., 2023), which enforces all-time constraint satisfaction. Among these methods, Tan and Dimarogonas (2021) and Wu et al. (2021) are originated from the right-hand side allocation strategy in Bertsekas (2016), and involve decomposing the coupling constraints with the help of additional variables and updating the additional variables along a direction such that the constraints stay feasible. Particularly, the authors in Tan and Dimarogonas (2021) studied problems with 1-dimensional coupling constraint, and presented a continuous-time distributed optimization algorithm based on finite-time consensus-seeking protocols. For problems subject to multi-dimensional coupling constraints, the authors in Wu et al. (2021, 2023) developed distributed feasible methods, where the update direction simultaneously decreases the objective function value and keeps the coupling constraints feasible. However, such update direction is found by each agent upon directly exchanging objective (Wu et al., 2021) or gradient information (Wu et al., 2023), which poses a potential risk of privacy leakage (Zhu, Liu, & Han, 2019). It is noteworthy that only convergence to a neighborhood of the optimal solution is guaranteed in Wu et al. (2021, 2023).

Some other notable methods achieving all-time constraint satisfaction include (Doostmohammadian et al., 2022; Turan & Alizadeh, 2022). The authors in Turan and Alizadeh (2022) considered a special class of constraint-coupled optimization problems, where the coefficients of the linear constraints belong to $\{0, 1\}$. A safety margin is constructed to avoid constraint violation by tightening the original constraint properly. In Doostmohammadian et al. (2022), 1-dimensional distributed optimization with equality constraint is studied. For this specific problem, the authors exploited the property that the local gradients evaluated over optimal solution are identical, and developed a distributed algorithm that averages the local gradients iteratively. The recursive satisfaction of coupling constraint follows from the property of Laplacian used for averaging. This idea was extended in Doostmohammadian (2023) and Doostmohammadian et al. (2023) to handle saturation and communication delays. The authors in Mestres and Cortés (2023) considered distributed optimization with nonlinear coupling constraints and introduced an equivalent reformulation with auxiliary variables that makes the overall problem separable. Then, a continuous-time algorithm was developed to solve the problem by cascading projected saddle-point dynamics with safe gradient flow.

To summarize, most distributed optimization algorithms handling coupling constraints in the literature cannot provide all-time constraint satisfaction, and those that can provide either focus on problems with specific structure such as one-dimensional variables and cannot be easily extended to handle a more general class of problems, or suffer from inexact convergence.

1.2. *Our contribution*

In this work, we consider networked optimization problems with separable convex objective functions and coupling multi-dimensional constraints in the form of both equalities and inequalities. Compared to existing constraint-coupled distributed optimization (Falsone et al., 2017; Notarnicola & Notarstefano, 2019; Wu et al., 2023), the proposed algorithms possess the following notable characteristics:

(i) They produce violation-free solutions whenever they are terminated, while also converging to precise solutions with an explicit rate guarantee.
(ii) They leverage the inherent structure of the coupling constraints, leading to enhanced communication efficiency and convergence performance.

To accomplish these desirable objectives, we reformulate multiple constraints by introducing auxiliary variables with a particular network-dependent linear transformation. This reformulation enables the decomposition of the problem, making it amenable to distributed solutions. The auxiliary variables corresponding to a certain constraint are introduced to an agent only if it is affected by the constraint, leading to a sparse and efficient formulation.

Subsequently, the reformulated problem is approached as a min-min optimization scenario, where the auxiliary and primal variables are optimized separately, and examined through sensitivity analysis. In particular, we show that the gradients of the objective function in the outer minimization are network-dependent affine transformations of Karush-Kuhn–Tucker (KKT) multipliers of the inner problem under mild conditions, and can be locally computed by agents. Provided that the local objective is strongly convex, we quantify the Lipschitz constants of the gradients, which facilitates the use of the accelerated dual averaging algorithm (Cohen, Diakonikolas, & Orecchia, 2018) in solving the reformulated problem. For general convex objectives, additional coordinate constraints are imposed on the auxiliary variables, which ensure the boundedness of the gradients. Based on this, the reformulated problem is solved by gradient descent with convergence guarantee.

Finally, one of the proposed algorithms is tested on a constrained consensus-seeking system under a control barrier function (CBF) based controller (Ames et al., 2016; Xu, Tabuada, Grizzle, & Ames, 2015), where a quadratic programming problem with sparse coupling constraints is solved at each sampling time. The results verify the effectiveness of the algorithm.

1.3. *Paper organization and notation*

The structure of this work is outlined as: Section 2 introduces the networked optimization problem with coupling constraints. In Section 3, we present a reformulation of the problem and propose a violation-free distributed optimization algorithm for strongly-convex problems, and provide convergence rate results. We extend the algorithm and analysis to general convex problems in Section 4. Section 5 presents the results of numerical experiments, and finally, Section 6 concludes this work.

$\mathbb{R}$, $\mathbb{R}^d$, and $\mathbb{R}^{n\times d}$ represent the 1-dimensional, d-dimensional, and $n \times d$ dimensional Euclidean spaces, respectively. $\mathbb{N}$ is the set of natural numbers. We denote by $\operatorname{mat}(A_1,\dots,A_n) = \begin{bmatrix} A_1 & \cdots & A_n \end{bmatrix}$, where $A_i, i = 1,\dots,n$ can be either a column vector or a matrix. $\operatorname{blkdiag}(g_1, g_2,\dots,g_n)$ denotes a block diagonal matrix with its diagonal blocks $g_1, g_2,\dots,g_n$, where $g_i, i = 1,\dots,n$ can be either a vector or a matrix, and $\operatorname{col}(h_1, h_2,\dots,h_n) = \left[h_1^T, h_2^T,\dots,h_n^T\right]^T$, where $h_i, i = 1,\dots,n$ is a column vector. Given an index set $\mathcal{E}$ and an integer M, we define the translation of an index set as $M + \mathcal{E} := \{x + M : x \in \mathcal{E}\}$.

2. Problem statement

2.1. Basic setup

Consider a class of linearly constrained optimization problems given by

$$\begin{aligned} &\min_{x_1,\dots,x_N} \sum_{i=1}^N f_i(x_i) \\ &\text{s.t.} \quad \sum_{i=1}^N A_i x_i + b_i \le 0, \\ &\qquad\ \sum_{i=1}^N E_i x_i + g_i = 0 \end{aligned} \tag{1}$$

where $x_i \in \mathbb{R}^{d_i}$, $d_i \in \mathbb{N}$, $A_i \in \mathbb{R}^{M\times d_i}$, $b_i \in \mathbb{R}^M$, $E_i \in \mathbb{R}^{Q\times d_i}$, $g_i \in \mathbb{R}^Q$, and $f_i : \mathbb{R}^{d_i} \to \mathbb{R}$ are local to each agent $i = 1,\dots N$. We denote an optimal solution to (1) by $x_i^*, i = 1,\dots,N$.

Each pair (A_i, b_i) (resp. (E_i, g_i)) has M (resp. Q) rows. Denote by $A_i^{[m]}$ the mth row of A_i and by $b_i^{[m]}$ the mth entry of b_i. Similarly, let $E_i^{[q]}$ and $g_i^{[q]}$ be the qth row and coordinate of E_i and g_i, respectively. Some of the rows can take zero values in the case of sparse coupling constraints. Hereafter, we denote by $\mathcal{V}^{[m]} = \{j : A_j^{[m]} \neq 0 \text{ or } b_j^{[m]} \neq 0\}$ (resp. $\mathcal{V}^{[M+q]} = \{j : E_j^{[q]} \neq 0 \text{ or } g_j^{[q]} \neq 0\}$) the set of agents that are influenced by the mth inequality constraint (resp. qth equality constraint), and by $\mathcal{I}_i$ (resp. $\mathcal{E}_i$) the set of indexes of inequality (resp. equality) constraints that influence agent i's decision variable.

2.2. Communication network

The communication among the agents is described by an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{S})$, where $\mathcal{V}$ is the set of agent indexes and $\mathcal{S}$ is the set of unordered pairs $(i, j), i, j \in \mathcal{V}$. Each pair $(i, j) \in \mathcal{S}$ represents a communication channel between i and j, and $\mathcal{S} \subseteq \mathcal{V} \times \mathcal{V}$ represents the set of communication channels. We denote by $\mathcal{N}_i$ the cluster of i's neighbors including itself, i.e., the set of agents that i can communicate with, i.e., $\mathcal{N}_i = \{j \in \mathcal{V} : (i, j) \in \mathcal{S}\} \cup \{i\}$.

For each scalar constraint indicated by the corresponding rows of A_i's and E_i's, the involved agents as well as the links between them in the communication $\mathcal{G}$ form an induced subgraph (Mesbahi & Egerstedt, 2010). We denote by $\mathcal{G}^{[m]} = (\mathcal{V}^{[m]}, \mathcal{S}^{[m]})$ (resp. $\mathcal{G}^{[M+q]} = (\mathcal{V}^{[M+q]}, \mathcal{S}^{[M+q]})$) the graph induced by the mth inequality (resp. qth equality) constraint. Specifically, $\mathcal{S}^{[m]} = \{(i, j) : (i, j) \in \mathcal{S}, i, j \in \mathcal{V}^{[m]}\}$ and the definition of $\mathcal{S}^{[M+q]}$ is similar. Accordingly, for each $\mathcal{G}^{[l]}, l = 1,\dots,M+Q$, we denote by $\mathcal{N}_i^{[l]}$ the cluster of i's neighbors including itself. The following example is presented to illustrate the communication network.

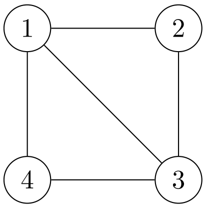


Fig. 1. A graph with 4 nodes.

Example. Consider problem (1) with $N = 4$, $d_i = 3, \forall i$, $M = Q = 1$,

$$\begin{pmatrix} A_1 \\ A_2 \\ A_3 \\ A_4 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix}, \qquad \begin{pmatrix} E_1 \\ E_2 \\ E_3 \\ E_4 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

and the communication graph in Fig. 1. Then,

$$\begin{aligned} &\mathcal{V}^{[1]} = \{1, 4\}, \quad \mathcal{S}^{[1]} = \{(1, 4)\}, \\ &\mathcal{V}^{[2]} = \{1, 2, 3\}, \quad \mathcal{S}^{[2]} = \{(1, 2), (2, 3), (1, 3)\}. \end{aligned}$$

The following assumption is made for the communication network.

Assumption 1. For each $l = 1,\dots,M+Q$, $\mathcal{G}^{[l]}$ is undirected and connected.

Assumption 1 is not restrictive in the sense that if all $A_i^{[m]}$, $m = 1,\dots,M$ and $E_i^{[M+q]}$, $q = 1,\dots,Q$ are non-zero then Assumption 1 holds when $\mathcal{G}^{[l]} = \mathcal{G}, l = 1,\dots,M+Q$ is connected.

Remark 1. We remark that we can let $\mathcal{G}^{[l]} = \mathcal{G}$ for simplicity, and we introduce $\mathcal{G}^{[l]}$, $l = 1,\dots,M+Q$ only to improve computation and communication efficiency and to relax the assumptions on A_i and E_i to ensure algorithm convergence. The introduction of subgraphs might affect the number of communication layers among agents, depending on the information exchange pattern. In particular, if each agent broadcasts certain information (to be specified in later sections) to its neighbors, then a single layer is needed and neighboring agents extract relevant information based on these induced subgraphs. Alternatively, if each agent exchanges information with other agents through one-to-one communication, then the induced subgraphs indicate the least communication channels needed. In other words, if agent i and agent j are neighbors in the graph $\mathcal{G}$ but not so in any induced subgraphs, then there is no need to establish a communication channel between them. The latter can be particularly effective when the coupling among agents, as indicated by the constraints, is sparse.

We associate each subgraph $\mathcal{G}^{[l]}$ with a *symmetric*, nonnegative weight matrix $P^{[l]} = [p_{ij}^{[l]}] \in \mathbb{R}^{|\mathcal{V}^{[l]}|\times|\mathcal{V}^{[l]}|}$, which will be used for the agents to weigh the exchanged information from their neighbors. The weights $p_{ij}^{[l]}$ satisfy the following conditions

$$p_{ij}^{[l]} = \begin{cases} >0, & j \in \mathcal{N}_i^{[l]} \\ 0, & \text{otherwise} \end{cases}, \qquad \sum_{j=1}^N p_{ij}^{[l]} = 1 \ \forall i. \tag{2}$$

Eq. (2) indicates the double stochasticity of $P^{[l]}$, and this requirement can be fulfilled by several well-known protocols, e.g., Metropolis–Hastings rule (Boyd, Diaconis, & Xiao, 2004). We note that if $\mathcal{G}^{[l]}$ is connected, there holds $\operatorname{Null}(I - P^{[l]}) = \operatorname{Span}(\mathbf{1})$ and $\operatorname{Range}(I - P^{[l]}) = \{z \in \mathbb{R}^{|\mathcal{V}^{[l]}|} : \mathbf{1}^T z = 0\}$, where $\operatorname{Null}(\cdot)$ and $\operatorname{Range}(\cdot)$ denotes the null space and the range of a linear map (Shi, Ling, Wu, & Yin, 2015).

Remark 2. The definition of subgraphs presents a trade-off between the conservativeness and effectiveness of the formulation. On one hand, Assumption 1 imposes a connectivity condition on each subgraph $\mathcal{G}^{[l]}$, which is a stronger condition than the connectivity of the graph $\mathcal{G}$. However, with this assumption, only the agents affected by the lth constraint will participate in addressing this constraint through communication and local computation, resulting in a more efficient algorithm in terms of computation and communication. On the other hand, if $\mathcal{G}^{[l]}$ is not connected, it is always possible to include more agents not affected by the lth constraint within $\mathcal{G}^{[l]}$ to satisfy Assumption 1, provided that $\mathcal{G}$ is connected. This increases the overall computation and communication load but makes the assumption less conservative.

3. Problem reformulation and algorithm

This section presents a novel reformulation of the optimization problem in (1). Based on sensitivity analysis, we show that the gradients of the objective function of the reformulated problem can be locally computed. Upon exploiting this feature, we develop an accelerated distributed optimization algorithm for solving (1).

3.1. Problem reformulation

Before proceeding to the reformulation, we introduce the following notation

$$\begin{aligned}
&\mathbf{A}^{[m]} = \mathrm{blkdiag}(\{A_l^{[m]}\}_{l\in\mathcal{V}^{[m]}}),\ \ b^{[m]} = \mathrm{col}(\{b_l^{[m]}\}_{l\in\mathcal{V}^{[m]}})\\
&\mathbf{E}^{[q]} = \mathrm{blkdiag}(\{E_l^{[q]}\}_{l\in\mathcal{V}^{[M+q]}}),\ \ g^{[q]} = \mathrm{col}(\{g_l^{[q]}\}_{l\in\mathcal{V}^{[M+q]}})\\
&x^{[m]} = \mathrm{col}(\{x_l\}_{l\in\mathcal{V}^{[m]}}), \qquad x^{[M+q]} = \mathrm{col}(\{x_l\}_{l\in\mathcal{V}^{[M+q]}}).
\end{aligned}$$

We incorporate a vector of slack variables to each constraint in (1), denoted by $y^{[l]} = \mathrm{col}(\{y_i^{[l]}\}_{i\in\mathcal{V}^{[l]}}) \in \mathbb{R}^{|\mathcal{V}^{[l]}|}, l = 1, \dots, M+Q$. Further denoting $y = \mathrm{col}(\{y^{[m]}\}_{m=1,\dots,M+Q})$ and $x = \mathrm{col}(\{x_i\}_{i=1,\dots,N})$, we arrive at the following new problem

$$\begin{aligned}
\min_{x,y}\ & \sum_{i=1}^{N} f_i(x_i)\\
\text{s.t.}\ \ & \mathbf{A}^{[m]}x^{[m]} + \left(I - P^{[m]}\right) y^{[m]} + b^{[m]} \le 0,\\
& \forall m = 1, \dots M\\
& \mathbf{E}^{[q]}x^{[M+q]} + \left(I - P^{[M+q]}\right) y^{[M+q]} + g^{[q]} = 0,\\
& \forall q = 1, \dots Q
\end{aligned} \tag{3}$$

where $P^{[m]}$ is the weight matrix defined in (2). The utilization of the linear mapping $(I - P^{[m]})$ for slack variables $y^{[m]}$ is motivated by two key factors. Firstly, it facilitates a network-aware decomposition of the problem, while ensuring the preservation of the solution for the original optimization problem (1), as outlined in Proposition 1. Secondly, compared to alternative linear mappings such as the graph Laplacian, $(I - P^{[m]})$ possesses a lower matrix norm. This characteristic is advantageous as it allows for the amplification of the step-size and therefore speeds up the convergence process.

Proposition 1. *Suppose Assumption 1 holds. The problems in (1) and (3) are equivalent in the sense that*

(i) they share the same objective function,
(ii) for any feasible solution (x, y) to (3), x is a feasible solution to (1),
(iii) for any feasible solution x to (1), there exist some y such that (x, y) is feasible to (3).

Proof. The statement in (i) holds trivially.

(ii) Given any feasible solution (x, y) to (3), it holds, for any m that

$$\begin{aligned}
&\mathbf{1}^T \left(\mathbf{A}^{[m]}x^{[m]} + \left(I - P^{[m]}\right) y^{[m]} + b^{[m]}\right)\\
&= \sum_{i\in\mathcal{V}} A_i^{[m]}x_i + b_i^{[m]} \le 0
\end{aligned}$$

and any q that

$$\begin{aligned}
&\mathbf{1}^T \left(\mathbf{E}^{[q]}x^{[M+q]} + \left(I - P^{[M+q]}\right) y^{[M+q]} + g^{[q]}\right)\\
&= \sum_{i\in\mathcal{V}} E_i^{[q]}x_i + g_i^{[q]} = 0
\end{aligned}$$

because of column stochasticity of $P^{[l]}, l = 1, \dots, M+Q$. Thus, x also satisfies the constraint in (1).

(iii) Given any feasible solution x to (1), we define

$$\begin{aligned}
r^{[m]} &= \mathbf{A}^{[m]}x^{[m]} + b^{[m]}, \forall m\\
v^{[q]} &= \mathbf{E}^{[q]}x^{[M+q]} + g^{[q]}, \forall q.
\end{aligned}$$

Then, it holds that $\bar{r}^{[m]} = \frac{\mathbf{1}^T r^{[m]}}{|\mathcal{V}^{[m]}|} \le 0$ and $\bar{v}^{[q]} = \frac{\mathbf{1}^T v^{[q]}}{|\mathcal{V}^{[M+q]}|} = 0$. Note that $\{(I - P^{[l]})y^{[l]} : y^{[l]} \in \mathbb{R}^{|\mathcal{V}^{[l]}|}\} = \mathrm{Range}(I - P^{[l]}) = \{z \in \mathbb{R}^{|\mathcal{V}^{[l]}|} : \mathbf{1}^T z = 0\}$. Thus, there exists a $y^{[l]}$ such that $(I - P^{[l]})y^{[l]} + r^{[l]} = \mathbf{1}\bar{r}^{[l]}$ for $l = 1, \dots, M$ and that $(I - P^{[l]})y^{[l]} + v^{[l]} = \mathbf{1}\bar{v}^{[l]}$ for $l = M+1, \dots, M+Q$. Based on this, a feasible solution (x, y) can be constructed for (3). ■

Given fixed slack variables $\{y^{[l]}\}_{l=1,\dots,M+Q}$, (3) can be partitioned and assigned to each agent. In particular, the local optimization problem for agent i is

$$\begin{aligned}
&\min_{x_i} f_i(x_i)\\
&\text{s.t. } A_i^{[m]}x_i + y_i^{[m]} - \sum_{j=1}^{N} p_{ij}^{[m]}y_j^{[m]} + b_i^{[m]} \le 0,\ \forall m \in \mathcal{I}_i\\
&\quad E_i^{[q]}x_i + y_i^{[M+q]} - \sum_{j=1}^{N} p_{ij}^{[M+q]}y_j^{[M+q]} + g_i^{[q]} = 0,\\
&\quad \forall q \in \mathcal{E}_i
\end{aligned} \tag{4}$$

where $\mathcal{I}_i$ and $\mathcal{E}_i$ denote the set of inequality constraints and the set of equality constraints affecting agent i, respectively.

The Lagrangian for problem (4) is

$$\begin{aligned}
&\mathcal{L}_i(x_i, \lambda_i, \mu_i) = f_i(x_i)\\
&+ \sum_{m\in\mathcal{I}_i} \left\langle \mu_i^{[m]}, A_i^{[m]}x_i + y_i^{[m]} - \sum_{j=1}^{N} p_{ij}^{[m]}y_j^{[m]} + b_i^{[m]} \right\rangle\\
&+ \sum_{q\in\mathcal{E}_i} \left\langle \lambda_i^{[q]}, E_i^{[q]}x_i + y_i^{[M+q]} - \sum_{j=1}^{N} p_{ij}^{[M+q]}y_j^{[M+q]} + g_i^{[q]} \right\rangle
\end{aligned} \tag{5}$$

where $\mu_i = \mathrm{col}(\{\mu_i^{[m]}\}_{m\in\mathcal{I}_i})$ and $\lambda_i = \mathrm{col}(\{\lambda_i^{[q]}\}_{q\in\mathcal{E}_i})$ are the multipliers of problem (4) corresponding to the inequality and equality constraints, respectively. We note that (4) is a linearly constrained optimization problem. Thus, for fixed y_i's, if $x_i(y_i)$ is an optimum of (4), then there exist multipliers $\mu_i(y_i)$ and $\lambda_i(y_i)$ such that the following KKT condition holds (Boyd & Vandenberghe, 2004)

$$\begin{aligned}
&\nabla_{x_i}\mathcal{L}_i\left(x_i(y_i), \lambda_i(y_i), \mu_i(y_i)\right) = 0\\
&\mu_i(y_i) \ge 0\\
&\left\langle \mu_i^{[m]}(y_i), A_i^{[m]}x_i + y_i^{[m]} - \sum_{j=1}^{N} p_{ij}^{[m]}y_j^{[m]} + b_i^{[m]} \right\rangle = 0,\\
&\forall m \in \mathcal{I}_i
\end{aligned} \tag{6}$$

Table 1
A list of main symbols in the problem reformulation.

M	Number of inequality constraints
Q	Number of equality constraints
$\mathcal{V}^{[m]}$	Cluster of agents affected by the mth inequality constraint
$\mathcal{V}^{[M+q]}$	Cluster of agents affected by the qth equality constraint
$\mathcal{I}_i$	Set of inequality constraints affecting agent i
$\mathcal{E}_i$	Set of equality constraints affecting agent i
$y_i^{[l]}$	Scalar variable associated with the lth constraint at agent i
$y^{[l]}$	$y^{[l]} = \mathrm{col}(\{y_i^{[l]}\}_{i\in\mathcal{V}^{[l]}})$
y_i	$y_i = \mathrm{col}\left(\{y_j^{[l]}\}_{j\in\mathcal{N}_i^{[l]},\forall l\in\mathcal{I}_i\cup(M+\mathcal{E}_i)}\right)$
y	$y = \mathrm{col}(\{y^{[l]}\}_{l=1,\ldots,M+Q})$

where

$$y_i = \mathrm{col}\left(\{y_j^{[l]}\}_{j\in\mathcal{N}_i^{[l]},\forall l\in\mathcal{I}_i\cup(M+\mathcal{E}_i)}\right)$$

is the collection of auxiliary variables that influence the local problem for agent i in (4).

We note that when solving (4) based on some common solvers, e.g., quadprog in MATLAB or CVX, the optimal primal and dual solutions can be simultaneously obtained.

Assumption 2. For each $i \in \mathcal{V}$, there holds

(i) f_i is ν-strongly convex, i.e., there is some $\nu \geq 0$ such that

$$f_i(y) - f_i(x) \geq \langle \nabla f_i(x), y - x\rangle + \frac{\nu}{2}\|y - x\|^2,$$

and has α-Lipschitz gradients, i.e., there exists some $\alpha > 0$ such that

$$\|\nabla f_i(y) - \nabla f_i(x)\| \leq \alpha\|y - x\|.$$

(ii) the matrix whose rows are $A_i^{[m]}$, $m \in \mathcal{I}_i$ and $E_i^{[q]}$, $q \in \mathcal{E}_i$ is full row rank.

Both Assumptions 1 and 2 (ii) hold under the following conditions: all $\mathrm{mat}(A_i^T, E_i^T)$'s have full column rank, and $\mathcal{G}$ is a connected network. These conditions naturally hold in, e.g., economic dispatch (Yang, Tan, & Xu, 2013).

Lemma 1. *Suppose Assumption 2 holds. Then, for the local constrained optimization problem in (4) with any given slack variables y_i,*

(i) feasiblility holds,
(ii) strong duality holds, and the KKT multipliers are unique.

Proof. Feasibility directly follows from Assumption 2-(ii). With feasibility, the refined Slater's condition automatically holds if the constraints are all linear (Boyd & Vandenberghe, 2004, Section 5.2.3), which together with Assumption 2-(i) yields strong duality. In addition, the vectors $\left\{A_i^{[m]}, m \in \mathcal{I}_i; E_i^{[q]}, q \in \mathcal{E}_i\right\}$ are linearly independent, known as the Linear Independence Constraint Qualification (LICQ), under which the KKT multipliers for the constrained optimization problem in (4) are guaranteed to be unique (Kyparisis, 1985). ■

For ease of reference, the main notations in the reformulated problem are summarized in Table 1.

3.2. Perturbed function and minimization algorithm

In this subsection, we view the problem in (4) as a perturbed version of the original problem (1), defined by

$$\begin{aligned}
\phi_i(y_i) &:= \min_{x_i} f_i(x_i)\\
\text{s.t.}\quad & A_i^{[m]}x_i + y_i^{[m]} - \sum_{j=1}^{N} p_{ij}^{[m]}y_j^{[m]} + b_i^{[m]} \leq 0, \forall m \in \mathcal{I}_i\\
& E_i^{[q]}x_i + y_i^{[M+q]} - \sum_{j=1}^{N} p_{ij}^{[M+q]}y_j^{[M+q]} + g_i^{[q]} = 0,\\
& \forall q \in \mathcal{E}_i.
\end{aligned}$$

Since (4) is always feasible under Assumption 2, $\phi_i(y_i)$ is well-defined. Using the definition of $\phi(y_i)$, the optimization problem in (3) can be equivalently expressed as

$$\min_{y}\left\{\phi(y) = \sum_{i=1}^{N}\phi_i(y_i)\right\}. \tag{7}$$

Upon substituting the optimal perturbation y^* into (3), the original problem (1) can be solved.

Remark 3. The rank condition in Assumption 2 can be relaxed. Indeed, in the case where Assumption 2-(ii) fails, one can consider a relaxed version of the perturbed problem:

$$\begin{aligned}
\phi_i'(y_i) &:= \min_{x_i,\rho_i\geq 0} f_i(x_i) + \omega\rho_i,\\
\text{s.t.}\quad & A_i^{[m]}x_i + y_i^{[m]} - \sum_{j=1}^{N} p_{ij}^{[m]}y_j^{[m]} + b_i^{[m]} \leq \rho_i, \forall m \in \mathcal{I}_i\\
& E_i^{[q]}x_i + y_i^{[M+q]} - \sum_{j=1}^{N} p_{ij}^{[M+q]}y_j^{[M+q]} + g_i^{[q]} = 0,\\
& \forall q \in \mathcal{E}_i
\end{aligned}$$

where ω is a positive scalar. If (1) is feasible and ω is sufficiently large, then there must exist some y_i such that x_i^* and $\rho_i = 0$, $i = 1, \ldots, N$ are optimal to the perturbed problem $\min_y \sum_{i=1}^{n}\phi_i'(y_i)$. Thus, by solving $\min_y \sum_{i=1}^{n}\phi_i'(y_i)$, one attains an optimal solution to (1) (Notarnicola & Notarstefano, 2019, Proposition III.3). Therefore, the algorithms presented next also apply to this setup, with minor modifications on the subproblem. However, for this relaxed problem, the projection of its feasible set onto the space of x_i is not identical to that of the original problem, i.e., the iterates during implementation are no longer guaranteed to be feasible to (1).

It is noteworthy that there is no consensus constraint when minimizing $\phi(y)$. As a result, further decomposition of this problem is not required, which is a key difference between the proposed method with existing literature (Notarnicola & Notarstefano, 2019). Next, we provide explicit expressions for the gradients of $\phi(y)$ in the following lemma.

Lemma 2. *Suppose Assumptions 1 and 2 hold with $\nu > 0$. There holds*

(i) $\phi(y)$ is convex and differentiable,
(ii) the gradients of $\phi(y)$ can be computed as

$$\begin{aligned}
\nabla_{y^{[m]}}\phi(y) &= (I - P^{[m]})^T\mathrm{col}(\{\mu_i^{[m]}(y_i)\}_{i\in\mathcal{V}^{[m]}})\\
\nabla_{y^{[M+q]}}\phi(y) &= (I - P^{[M+q]})^T\mathrm{col}(\{\lambda_i^{[q]}(y_i)\}_{i\in\mathcal{V}^{[M+q]}})
\end{aligned} \tag{8}$$

where $\mu_i^{[m]}(y_i)$ and $\lambda_i^{[q]}(y_i)$ denote the KKT multipliers associated with the corresponding inequality and equality constraints in (4),

(iii) $\phi(y)$ has α_ϕ-Lipschitz continuous gradients for some finite α_ϕ.

Proof. (i) The convexity of ϕ can be proved by following Boyd and Vandenberghe (2004, Section 5.6.1). For completeness, a proof is included. Let x and x' be the optimal primal solutions of (7) under y and y'. Given any $\lambda \in [0, 1]$, $\phi(\lambda y + (1-\lambda)y')$ is well-defined, and there holds

$$\lambda\phi(y) + (1-\lambda)\phi(y')$$
$$= \lambda \sum_{i=1}^{N} f_i(x) + (1-\lambda)\sum_{i=1}^{N} f_i(x')$$
$$\geq \sum_{i=1}^{N} f_i(\lambda x + (1-\lambda)x')$$

where the inequality is due to the convexity of f_i. In addition, $\lambda x + (1-\lambda)x'$ is feasible to the optimization problem corresponding to $\phi(\lambda y + (1-\lambda)y')$ due to linearity of the constraints. Thus, by optimality, we obtain

$$\lambda\phi(y) + (1-\lambda)\phi(y') \geq \phi(\lambda y + (1-\lambda)y').$$

According to Lemma 1, the KKT multipliers are unique for any auxiliary variable y, and therefore $\phi(y)$ is globally differentiable (Florenzano & Le Van, 2001, Corollary 7.3.1).

(ii) Because strong duality holds, we obtain from sensitivity analysis (Boyd & Vandenberghe, 2004, Section 5.6.2) that

$$\phi_i(y_i')$$
$$\geq \phi_i(y_i) + \sum_{m\in\mathcal{I}_i}\left\langle \mu_i^{[m]}(y_i), (I-P^{[m]})_i\left((y')^{[m]} - y^{[m]}\right)\right\rangle$$
$$+ \sum_{q\in\mathcal{E}_i}\left\langle \lambda_i^{[q]}(y_i), (I-P^{[M+q]})_i\left((y')^{[M+q]} - y^{[M+q]}\right)\right\rangle$$
$$= \phi_i(y_i) + \sum_{m\in\mathcal{I}_i}\left\langle (I-P^{[m]})_i^T \mu_i^{[m]}(y_i), (y')^{[m]} - y^{[m]}\right\rangle$$
$$+ \sum_{q\in\mathcal{E}_i}\left\langle (I-P^{[M+q]})_i^T \lambda_i^{[q]}(y_i), (y')^{[M+q]} - y^{[M+q]}\right\rangle$$

where $(\cdot)_i$ represents the ith row of a matrix. This indicates that

$$\nabla_{y^{[m]}}\phi_i(y_i) = (I-P^{[m]})_i^T \mu_i^{[m]}(y_i), \forall m \in \mathcal{I}_i$$

and

$$\nabla_{y^{[M+q]}}\phi_i(y_i) = (I-P^{[M+q]})_i^T \lambda_i^{[q]}(y_i), \forall q \in \mathcal{E}_i$$

where $\nabla_{y^{[l]}}\phi_i$ denotes the block coordinate gradient with respect to $y^{[l]}$. Note that $\phi(y) = \sum_{i=1}^N \phi_i(y_i)$ and each y_i is partially coupled with each other. For each coordinate block indexed by m, $\nabla_{y^{[m]}}\phi(y)$ is contributed only by $\mu_i^{[m]}(y_i)(I-P^{[m]})_i^T, i \in \mathcal{V}^{[m]}$, that is, $\forall m = 1, \ldots, M$

$$\nabla_{y^{[m]}}\phi(y) = \sum_{i\in\mathcal{V}^{[m]}} \mu_i^{[m]}(y_i)(I-P^{[m]})_i^T$$
$$= (I-P^{[m]})^T \mathrm{col}(\{\mu_i^{[m]}(y_i)\}_{i\in\mathcal{V}^{[m]}})$$

and, similarly, $\forall q = 1, \ldots, Q$

$$\nabla_{y^{[M+q]}}\phi(y) = (I-P^{[M+q]})^T \mathrm{col}(\{\lambda_i^{[q]}(y_i)\}_{i\in\mathcal{V}^{[M+q]}}). \tag{9}$$

(iii) Because each f_i is strongly convex, the second-order sufficient conditions (SOSC) automatically hold, which implies that the KKT multipliers are Lipschitz continuous with finite positive parameters (Subotić, Hauswirth, & Dörfler, 2021). Based on the formulas of gradients given in (ii), we obtain that ϕ has Lipschitz continuous gradients with a finite positive parameter. ■

Algorithm 1 Violation-free distributed accelerated dual averaging

```
1: input: arbitrary variable $y$, parameters $\{\gamma_t\}_{t\geq 0}$ and $\{\Gamma_t = \sum_{\tau=1}^{t}\gamma_\tau\}_{t\geq 0}$
2: output: $x_i(\hat{y}_i), i = 1, \ldots, n$
3: for $t = 1, 2, \ldots$ do
4:   for each agent $i \in \{1, \ldots, N\}$:
5:   collect $y_j^{[l]}$ from each $j \in \mathcal{N}_i^{[l]}, \forall l \in \mathcal{I}_i \cup (M + \mathcal{E}_i)$
6:   compute $x_i(y_i)$ and the multipliers $\mu_i(y_i)$ and $\lambda_i(y_i)$ by solving (4)
7:   collect $\mu_j^{[m]}(y_j)$ from each $j \in \mathcal{N}_i^{[m]}, \forall m \in \mathcal{I}_i$ and $\lambda_i^{[q]}(y_i)$ from each $j \in \mathcal{N}_i^{[M+q]}, \forall q \in \mathcal{E}_i$
8:   compute $\nabla_{y_i^{[l]}}\phi(y), \forall l \in \mathcal{I}_i \cup (M + \mathcal{E}_i)$ according to (8)
9:   if $t = 1$ then
10:    set $\hat{y}_i^{[l]} = z_i^{[l]} = -\gamma_1 \nabla_{y_i^{[l]}}\phi(y), \forall l \in \mathcal{I}_i \cup (M + \mathcal{E}_i)$
11:  else
12:    update $y_i^{[l]}, \forall l \in \mathcal{I}_i \cup (M + \mathcal{E}_i)$ by
         $y_i^{[l]} \leftarrow \left(1 - \frac{\gamma_t}{\Gamma_t}\right)\hat{y}_i^{[l]} + \frac{\gamma_t}{\Gamma_t} z_i^{[l]}$
         $z_i^{[l]} \leftarrow z_i^{[l]} - \gamma_t \nabla_{y_i^{[l]}}\phi(y)$
         $\hat{y}_i^{[l]} \leftarrow \left(1 - \frac{\gamma_t}{\Gamma_t}\right)\hat{y}_i^{[l]} + \frac{\gamma_t}{\Gamma_t} z_i^{[l]}$
13:  end if
14: end for
```

We have converted the original problem represented in (1) into the form (7). Consequently, finding a solution for (1) is equivalent to solving (7). To accomplish this, it becomes essential to compute the block coordinate gradient of $\nabla_{y^{[l]}}\phi(y)$ in a distributed manner. Lemma 2 indicates that each agent $i \in \mathcal{V}^{[l]}$ can collaboratively compute $\nabla_{y^{[l]}}\phi(y)$, $l = 1, \ldots, M + Q$ by communicating with their immediate neighbors about the KKT multipliers. This is partially due to the fact that the closed-form expression of the block coordinate gradient involves a linear mapping $(I - P^{[m]})^T$, which has been intentionally designed to be compatible with the communication network. Based on this fact, the accelerated dual averaging method (Cohen et al., 2018) can be implemented to solve (7) in a distributed manner.

The overall algorithm is summarized in Algorithm 1. Each agent updates two sequences of auxiliary variables: y_i and $\hat{y}_i$, according to the following rules. In step 5, each agent communicates with its neighbors to collect y_j variables to formulate a local problem in the form of (4). After locally identifying the Lagrangian multipliers for (4), each agent exchanges the multipliers with its neighbors in order to compute the block coordinate gradient of $\phi(y)$, as outlined in steps 6–8. Finally, each agent follows the accelerated dual averaging to update y_i in step 12. We remark that the parameters γ_t and Γ_t should be properly chosen to accelerate the convergence (Cohen et al., 2018), as outlined in Theorem 1.

4. Convergence analysis

4.1. Computing lipschitz constants

Before establishing conditions under which Algorithm 1 converges, we quantify the Lipschitz constant α_ϕ of the gradients of $\phi(y)$. In particular, we present the Lipschitz constants of $\mu_i(y_i)$ and $\lambda_i(y_i)$, based on which the Lipschitz constant of $\nabla\phi(y)$ can be computed.

Under Assumption 2 with $\nu > 0$, (3) has a unique regular optimizer for any $y^{[l]}, l = 1, \ldots, M+Q$. In addition, the KKT multipliers are globally continuous with respect to y (Subotić et al., 2021, Theorem 2).

Proposition 2. *Suppose Assumptions 1 and 2 hold with $\nu > 0$. The Lipschitz constant of the gradients of $\phi(y)$ can be computed as*

$$\alpha_\phi \le \left(\max_i \alpha_{(\lambda_i,\mu_i)}\right)\left(\max_{l\in\{1,\ldots,M+Q\}} \|I - P^{[l]}\|\sqrt{|\mathcal{V}^{[l]}|}\right)\sqrt{M+Q} \tag{10}$$

where $\alpha_{(\lambda_i,\mu_i)}$ denotes the Lipschitz constant of $\mu_i(y_i)$ and $\lambda_i(y_i)$, given by

$$\alpha_{(\lambda_i,\mu_i)} \le \left(\max_{l\in\mathcal{I}_i\cup(M+\mathcal{E}_i)} \|I - P^{[l]}\|\right)\sqrt{\frac{\alpha}{\lambda_{\min}(\hat{\mathcal{B}}_i^T\hat{\mathcal{B}}_i)}}$$

where α denotes the Lipschitz constant of ∇f_i, $\hat{\mathcal{B}}_i = \mathrm{mat}(\hat{E}_i^T, \hat{A}_i^T)$, and $\mathrm{mat}(A, B) = \big[A, B\big]$.

Proof. Recall the Lagrangian in (5) and consider the KKT system

$$\mathcal{F}_i(x_i, \lambda_i, \mu_i, y_i) := \left[\begin{array}{c} \nabla_{x_i}\mathcal{L}_i(x_i, \lambda_i, \mu_i) \\ \hline \hat{E}_i x_i + \hat{g}_i + \mathrm{blkdiag}(\{(I - P^{[M+q]})_i\}_{q\in\mathcal{E}_i})y_i^{\mathcal{E}_i} \\ \hline \mathrm{diag}(\{\mu_i^{[m]}\}_{m\in\mathcal{I}_i})\times \\ \left(\hat{A}_i x_i + \hat{b}_i + \mathrm{blkdiag}(\{(I - P^{[m]})_i\}_{m\in\mathcal{I}_i})y_i^{\mathcal{I}_i}\right) \end{array}\right] = 0$$

where

$$y_i^{\mathcal{I}_i} = \mathrm{col}(\{y^{[m]}\}_{m\in\mathcal{I}_i}) \text{ and } y_i^{\mathcal{E}_i} = \mathrm{col}(\{y^{[M+q]}\}_{q\in\mathcal{E}_i}),$$

and

$$\hat{E}_i = [E_i^{[1]}; \ldots; E_i^{|\mathcal{E}_i|}] \text{ and } \hat{g}_i = \mathrm{col}(\{g_i^{[q]}\}_{q\in\mathcal{E}_i}),$$

$$\hat{A}_i = [A_i^{[1]}; \ldots; A_i^{|\mathcal{I}_i|}] \text{ and } \hat{b}_i = \mathrm{col}(\{b_i^{[m]}\}_{m\in\mathcal{I}_i}).$$

Next, to ease notation, we omit the argument from any map that solely depends on y.

We bound the Lipschitz constants for functions $\mu_i^{[m]}$ and $\lambda_i^{[q]}$ by following the approach in Subotić et al. (2021, Proposition 4). Denote by

$$\mathbf{I}_i := \left\{ m : A_i^{[m]} x_i + y_i^{[m]} - \sum_{j\in\mathcal{N}_i^{[m]}} p_{ij}^{[m]} y_j^{[m]} + b_i^{[m]} = 0 \right\}$$

the set of active inequality constraints for agent i, and by $\bar{\mathbf{I}}_i$ the set of inactive constraints for i. Without loss of generality, we assume the first $|\mathbf{I}_i|$ inequality constraints are active. For simplicity, the inequality constraints are required to be strictly active, that is, the KKT multiplier associated with the constraint should be strictly positive. As a consequence, the KKT multipliers are differentiable with respect to y_i (Subotić et al., 2021). However, such bound can be generalized to the case with weakly active inequality constraints by following Jittorntrum (1978).

By the implicit function theorem (Bertsekas, 2016, Appendix A) and by organizing the matrices according to $(x_i, [\lambda_i, \mu_i^{\mathbf{I}_i}], \mu_i^{\bar{\mathbf{I}}_i})$, it holds that

$$\nabla_{y_i}(x_i, \lambda_i, \mu_i) = -(\nabla_{x_i,\lambda_i,\mu_i}\mathcal{F}_i)^{-1}\nabla_{y_i}\mathcal{F}_i \tag{11}$$

where

$$\nabla_{x_i,\lambda_i,\mu_i}\mathcal{F}_i := \begin{bmatrix} \nabla^2_{x_ix_i}\mathcal{L}_i & \mathcal{B}_i & \hat{A}_i^{\bar{\mathbf{I}}_i} \\ \mathcal{D}_{\mathbf{I}_i}\mathcal{B}_i^T & 0 & 0 \\ 0 & 0 & \mathrm{diag}(\mathcal{H}_i^{\bar{\mathbf{I}}_i}) \end{bmatrix},$$

$$\nabla_{y_i}\mathcal{F}_i := \begin{bmatrix} 0 \\ \mathcal{D}_{\mathbf{I}_i}\mathcal{G}_i \\ 0 \end{bmatrix}, \quad \mathcal{B}_i = \mathrm{mat}(\hat{E}_i^T, \hat{A}_i^{\mathbf{I}_i}),$$

$$\mathcal{D}_{\mathbf{I}_i} = \mathrm{diag}([\mathbf{1}_{|\mathcal{E}_i|}; \mathrm{col}(\{\mu_i^{[m]}\}_{m\in\mathbf{I}_i})]),$$

$$\hat{A}_i^{\mathbf{I}_i} = \mathrm{mat}(\{(A_i^{[m]})^T\}_{m\in\mathbf{I}_i}), \quad \hat{A}_i^{\bar{\mathbf{I}}_i} = \mathrm{mat}(\{(A_i^{[m]})^T\}_{m\in\bar{\mathbf{I}}_i}),$$

$$\mathcal{G}_i = \mathrm{blkdiag}(\{(I - P^{[M+q]})_i\}_{q\in\mathcal{E}_i}, \{(I - P^{[m]})_i\}_{m\in\mathbf{I}_i}),$$

$$\mathcal{H}_i = \hat{A}_i x_i + \hat{b}_i + \mathrm{blkdiag}(\{(I - P^{[m]})_i\}_{m\in\mathcal{I}_i})y_i^{\mathcal{I}_i},$$

$\mathcal{H}_i^{\bar{\mathbf{I}}_i} = \mathrm{col}(\{\mathcal{H}_i^{[m]}\}_{m\in\bar{\mathbf{I}}_i})$, and $\nabla^2_{x_ix_i}$ represents the Hessian operator. It can be verified that

$$(\nabla_{x_i,\lambda_i,\mu_i}\mathcal{F}_i)^{-1} = \begin{bmatrix} \mathcal{M}_i^{-1} & * \\ 0 & * \end{bmatrix} \text{ with } \mathcal{M}_i = \begin{bmatrix} \nabla^2_{x_ix_i}\mathcal{L}_i & \mathcal{B}_i \\ \mathcal{D}_{\mathbf{I}_i}\mathcal{B}_i^T & 0 \end{bmatrix}$$

where the symbol $*$ represents non-zero components that are irrelevant to the context. Therefore, for the inactive inequality constraints, it holds that $\nabla_{y_i}\mu_{\bar{\mathbf{I}}_i} = 0$. And for equality and active inequality constraints, we obtain from the formula of the inverse of 2 × 2 block matrices (Lu & Shiou, 2002) that

$$\nabla_{y_i}\begin{bmatrix} \lambda_i \\ \mu_i^{\mathbf{I}_i} \end{bmatrix} = (\mathcal{M}_i^{-1})_{22}\mathcal{D}_{\mathbf{I}_i}\mathcal{G}_i = (\mathcal{B}_i^T(\nabla^2_{xx}\mathcal{L}_i)^{-1}\mathcal{B}_i)^{-1}\mathcal{G}_i$$

where $(\mathcal{M}_i^{-1})_{22}$ denotes the $(2, 2)$th entry of $\mathcal{M}_i^{-1}$. This indicates that the Lipschitz constants of μ_i and λ_i

$$\begin{aligned} \alpha_{(\lambda_i,\mu_i)} &\le \|(\mathcal{B}_i^T(\nabla^2_{xx}\mathcal{L}_i)^{-1}\mathcal{B}_i)^{-1}\mathcal{G}_i\| \\ &\le \|(\mathcal{B}_i^T(\nabla^2_{xx}\mathcal{L}_i)^{-1}\mathcal{B}_i)^{-1}\|\|\mathcal{G}_i\| \\ &\le \|\mathcal{G}_i\|\sqrt{\frac{\alpha}{\lambda_{\min}(\hat{\mathcal{B}}_i^T\hat{\mathcal{B}}_i)}} \\ &\le \left(\max_{l\in\mathcal{I}_i\cup(M+\mathcal{E}_i)} \|I - P^{[l]}\|\right)\sqrt{\frac{\alpha}{\lambda_{\min}(\hat{\mathcal{B}}_i^T\hat{\mathcal{B}}_i)}} \end{aligned}$$

where the third and fourth inequality follows from

$$\|\mathcal{B}_i^T(\nabla^2_{xx}\mathcal{L}_i)^{-1}\mathcal{B}_i\| \ge \sqrt{\frac{\lambda_{\min}(\mathcal{B}_i^T\mathcal{B}_i)}{\alpha}} \ge \sqrt{\frac{\lambda_{\min}(\hat{\mathcal{B}}_i^T\hat{\mathcal{B}}_i)}{\alpha}}$$

and

$$\|\mathrm{blkdiag}(A, B)\| = \max\{\|A\|, \|B\|\} \text{ and } \|(A)_i\| \le \|A\|,$$

respectively. We note that the above bound of $\alpha_{(\lambda_i,\mu_i)}$ holds for uncertain $\mathbf{I}_i$.

Given y and y', we have

$$\begin{aligned} &\|\mathrm{col}(\{\mu_i^{[m]}(y_i)\}_{i\in\mathcal{V}^{[m]}}) - \mathrm{col}(\{\mu_i^{[m]}(y_i')\}_{i\in\mathcal{V}^{[m]}})\| \\ &= \sqrt{\sum_{i\in\mathcal{V}^{[m]}} \|\mu_i^{[m]}(y_i) - \mu_i^{[m]}(y_i')\|^2} \\ &\le (\max_i \alpha_{(\lambda_i,\mu_i)})\sqrt{\sum_{i\in\mathcal{V}^{[m]}} \|y_i - y_i'\|^2} \\ &= (\max_i \alpha_{(\lambda_i,\mu_i)})\sqrt{|\mathcal{V}^{[m]}|}\|y - y'\|. \end{aligned}$$

This in conjunction with Lemma 2 leads to

$$\begin{aligned} &\|\nabla_{y^{[m]}}\phi(y) - \nabla_{y^{[m]}}\phi(y')\| \\ &\le \|I - P^{[m]}\|(\max_i \alpha_{(\lambda_i,\mu_i)})\sqrt{|\mathcal{V}^{[m]}|}\|y - y'\|. \end{aligned}$$

Furthermore,

$$
\begin{aligned}
&\|\nabla_y\phi(y) - \nabla_y\phi(y')\| \\
&= \sqrt{\sum_{l=1}^{M+Q} \|\nabla_{y^{[l]}}\phi(y) - \nabla_{y^{[l]}}\phi(y')\|^2} \\
&\le (\max_i \alpha_{(\lambda_i,\mu_i)}) \left(\max_{l\in\{1,\ldots,M+Q\}} \|I - P^{[l]}\| \sqrt{|\mathcal{V}^{[l]}|} \right) \\
&\quad \times \sqrt{M+Q}\,\|y - y'\|.
\end{aligned}
$$

The proof is complete. ■

4.2. Rate of convergence

Theorem 1. *Suppose the premise given in Lemma 2 holds. If the parameter γ_t in Algorithm 1 is set as $\gamma(t+1)$ where $\gamma \le 1/(2\alpha_\phi)$, then, for all $t \ge 2$, the solution to the problem in (3) with auxiliary variable $y = \hat{y}^{(t)}$ is a feasible solution to (1) and satisfies*

$$
\sum_{i=1}^{N} f_i(x_i(\hat{y}_i^{(t)})) - f_i(x_i^*) \le \frac{\|y^{(0)} - y^*\|^2}{2\Gamma_t} = \frac{2\alpha_\phi\|y^{(0)} - y^*\|^2}{t(t+3)}
$$

where $y^{(0)}$ denotes the initial variable and y^* is any (fixed) minimizer of (7).

Proof. The proof is a consequence of Cohen et al. (2018, Theorem 3.4) and omitted here for brevity. ■

5. Extension to non-strongly convex case

In this section, we relax the strong convexity assumption on f_i.

For general convex f_i, following the same proof of Lemma 1 $\phi(y)$ remains differentiable. However, the KKT multipliers as well as the gradients of $\phi(y)$ are not necessarily Lipschitz continuous, making Algorithm 1 not applicable. In addition, the minimization problem in (7) is unconstrained and has noncompact solution set. To guarantee the boundedness of $\nabla\phi(y)$, we introduce a compact constraint

$$
\mathcal{Y} := [-C, C]^{\sum_{l=1}^{M+Q}|\mathcal{V}^{[l]}|}
$$

to y, where C is a positive constant. Note that $\mathcal{Y}$ imposes constraint on each individual coordinate of y, which facilitates projection in local updates of $y_i^{[m]}$. If C is sufficiently large, this constraint will not change the optimal solution of the problem in (7). Upon incorporating the constraints, we arrive at the following optimization problem:

$$
\min_{y\in\mathcal{Y}} \phi(y). \tag{12}
$$

Proposition 3. *Let y^* be an optimal solution of problem (7) and C is a positive constant such that $y^* \in \mathcal{Y}$. Then, y^* is also an optimal solution of problem (12) and the two problems have identical optimal cost.*

Proof. Since problem in (7) is unconstrained, it has lower or identical optimal cost than that in (12). Because $y^* \in \mathcal{Y}$, they lead to a cost of (12) that is equal to the optimal cost (7), implying that y^* is also an optimal solution of (12) and the two problems have identical optimal cost. ■

The compactness of $\mathcal{Y}$, together with the convexity of $\phi(y)$, leads to the following lemma, whose proof can be found in Bertsekas (2016, Appendix B).

Algorithm 2 Violation-free distributed gradient method

```
input: arbitrary variable $y$, and parameter $\{\gamma_t\}_{t\ge0}$
output: $x_i(y_i), i = 1, \ldots, n$
for $t = 1, 2, \ldots$ do
    for each agent $i \in \{1, \ldots, N\}$:
    collect $y_j^{[l]}$ from each $j \in \mathcal{N}_i^{[l]}, \forall l \in \mathcal{I}_i \cup (M + \mathcal{E}_i)$
    compute $x_i(y_i)$ and the multipliers $\mu_i(y_i)$ and $\lambda_i(y_i)$ by solving (4)
    collect $\mu_j^{[m]}(y_j)$ from each $j \in \mathcal{N}_i^{[m]}, \forall m \in \mathcal{I}_i$, and $\lambda_i^{[q]}(y_i)$ from each $j \in \mathcal{N}_i^{[M+q]}, \forall q \in \mathcal{E}_i$
    compute $\nabla_{y_i^{[l]}}\phi(y), \forall l \in \mathcal{I}_i \cup (M + \mathcal{E}_i)$ according to (8)
    update $y_i^{[l]}, \forall l \in \mathcal{I}_i \cup (M + \mathcal{E}_i)$ by
        $y_i^{[l]} \leftarrow \mathrm{Proj}_{[-C,C]}(y_i^{[l]} - \gamma_t \nabla_{y_i^{[l]}}\phi(y))$
end for
```

Lemma 3. *Under Assumption 2, the gradients of $\phi(y)$ in (12) are bounded from above, i.e., $\|\nabla\phi(y)\| \le G$.*

We note that in the non-strongly convex case, the acceleration technique in Algorithm 1 may not be applicable, because the objective function $\phi(y)$ defined in (7) may be non-smooth. Nevertheless, based on Lemma 3, standard gradient descent with diminishing step-size can be used to solve (12) with convergence guarantees (Beck, 2017, Theorem 8.30). The algorithm is summarized in Algorithm 2.

Theorem 2. *Suppose Assumptions 1 and 2 hold with $\nu = 0$. If the parameter γ_t in Algorithm 2 is set as $\frac{\sqrt{2\Theta}}{G\sqrt{t+1}}$ where $\Theta = \max_{x,y\in\mathcal{Y}} \|x - y\|^2/2$, then, for all $t \ge 2$, the solution to the problem in (3) is a feasible solution to (1) and satisfies*

$$
\min_{\tau=0,1,\ldots,t} \sum_{i=1}^{N} f_i(x_i(y_i^{(\tau)})) - f_i(x_i^*) \le \frac{2(1+\log(3))G\sqrt{2\Theta}}{\sqrt{t+2}}.
$$

6. Numerical experiment

In this section, we apply the proposed algorithm to solve a multi-agent consensus problem subject to two coupling state constraints.

6.1. Problem setup

Consider a multi-agent system with $N = 7$ agents. They are connected via a line graph. For each agent i, the state is denoted as $z_i \in \mathbb{R}^2$ and the system dynamics is $\dot{z}_i = x_i$. Define $z = \mathrm{col}(\{z_i\}_{i=1,\ldots,7})$ and $x = \mathrm{col}(\{x_i\}_{i=1,\ldots,7})$. The agents seek consensus under two coupling state constraints given by

$$
\begin{aligned}
\mathcal{Z} := \Big\{ & z \in \mathbb{R}^{14}, g_1(\mathrm{col}(\{z_i\}_{i=1,\ldots,4})) \ge 0, \\
& g_2(\mathrm{col}(\{z_i\}_{i=4,\ldots,7})) \ge 0 \Big\}
\end{aligned} \tag{13}
$$

where

$$
\begin{aligned}
g_1(\mathrm{col}(\{z_i\}_{i=1,\ldots,4})) &= 4 - \sum_{i=1}^{4} z_i^T z_i \\
g_2(\mathrm{col}(\{z_i\}_{i=4,\ldots,7})) &= 16 - \sum_{i=4}^{7} (z_i - [2;2])^T (z_i - [2;2]).
\end{aligned} \tag{14}
$$

To solve this constrained control problem, we use the consensus protocol $x_{nom,i} = \sum_{j \in \mathcal{N}_i}(z_j - z_i)$ and follow the CBF approach (Ames et al., 2016; Xu et al., 2015) to handle the constraints. To proceed, we present the definition of CBF particularly for the dynamical system

$$\dot{z} = x. \tag{15}$$

Definition 1 (*CBF*)**.** Consider the safety set $\mathcal{Z}$ and the dynamical system given by (13) and (15), respectively. The functions in (14) are CBFs if there exists a locally Lipschitz, strictly increasing function $\alpha(\cdot)$ with $\alpha(0) = 0$ such that $\forall z$, $\exists x$

$$\begin{aligned} &\left\langle \nabla g_1(\mathrm{col}(\{z_i\}_{i=1,\ldots,4})), \mathrm{col}(\{x_i\}_{i=1,\ldots,4}) \right\rangle \\ &+ \alpha(g_1(\mathrm{col}(\{z_i\}_{i=1,\ldots,4}))) \geq 0, \end{aligned} \tag{16}$$

and

$$\begin{aligned} &\left\langle \nabla g_2(\mathrm{col}(\{z_i\}_{i=4,\ldots,7})), \mathrm{col}(\{x_i\}_{i=4,\ldots,7}) \right\rangle \\ &+ \alpha(g_2(\mathrm{col}(\{z_i\}_{i=4,\ldots,7}))) \geq 0. \end{aligned} \tag{17}$$

It has been shown that any locally Lipschitz x that fulfils the CBF constraint in (16) and (17) makes the set $\mathcal{Z}$ forward invariant, and if $\mathcal{Z}$ is compact, asymptotically stable (Xu et al., 2015). Following this fact and setting α as the identity function, for any given states z, an optimization-based controller can be constructed as

$$\begin{aligned} \min_{x_1,\ldots,x_N} \ & \sum_{i=1}^{N} \frac{1}{2} \|x_i - x_{nom,i}\|^2 \\ \text{s.t.} \ & \text{constraints in (16) and (17).} \end{aligned} \tag{18}$$

6.2. Performance on a problem instance

In this subsection, we take the optimization problem instance with $z_i = z_i(0) = (\cos(2\pi i/7) + 2, \sin(2\pi i/7) + 1), i = 1, \ldots, N$. to examine the performance of Algorithm 1. For each constraint $m = 1, 2$, we take

$$P^{[m]} = \begin{bmatrix} 2/3 & 1/3 & 0 & 0 \\ 1/3 & 1/3 & 1/3 & 0 \\ 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 1/3 & 2/3 \end{bmatrix}$$

which is compatible with the line graph among the agents. Both Assumptions 1 and 2 are met in this problem.

We compare Algorithm 1 with two existing algorithms applicable to this problem: the dual subgradient method in Falsone et al. (2017) and the buffering drift-plus-penalty algorithm (referred to as B-DPP) (Wang, Zhu, Ou and Lu, 2023). For Falsone et al. (2017), we choose the step-size parameter $1/(t+1)$. For Wang, Zhu et al. (2023), we set the parameters according to Theorem 1 therein. The dual variables for both of them are initialized to zero. In addition, we incorporate the sparse formulation from Section 2 into them to save communication and computation resources for a fair comparison. For Algorithm 1, the auxiliary variables are initialized to zero. We calculate and use the upper bound of $\gamma = 0.346$ according to the condition given in Theorem 1. We use the quadratic programming solver in MATLAB to solve the global problem (18) for benchmarking and also the local subproblems in Algorithm 1.

The trajectories of the primal objective error, the primal variable error, and the values of the two coupling constraint functions are plotted in Figs. 2–5, respectively. We observe from Figs. 2 and 3 that Algorithm 1 outperforms (Falsone et al., 2017; Wang, Zhu et al., 2023) in terms of convergence rate. We remark that the oscillation phenomenon is common in accelerated algorithms. Indeed, this is reflected in the theoretical analysis, where a Lyapunov function, different from the objective error, is shown to decrease monotonically (see the proof of Theorem 3.4 in Cohen et al., 2018). Additionally, Figs. 4 and 5 demonstrate that the two coupling constraints are satisfied at all times by Algorithm 1, whereas the dual subgradient method and B-DPP violate the first constraint in the early phase of execution.

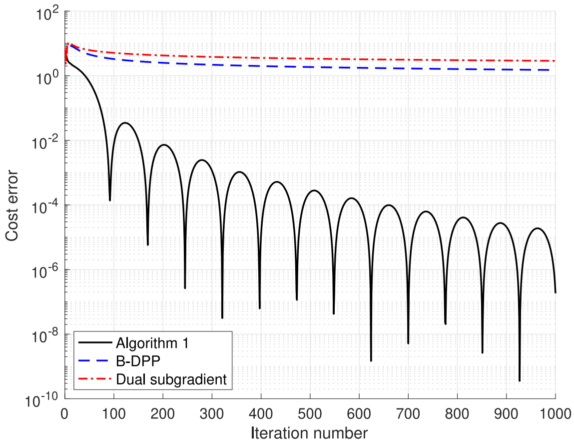


Fig. 2. Convergence of primal objective error $\sum_{i=1}^{N}(f_i(x_i) - f_i(x_i^*))$.

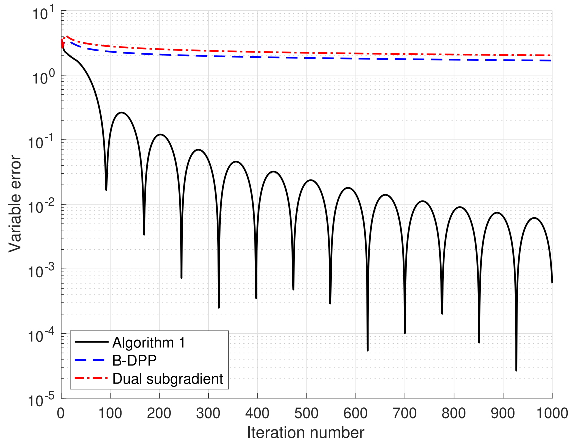


Fig. 3. Convergence of primal variable error $\sqrt{\sum_{i=1}^{N} \|x_i - x_i^*\|^2}$.

Algorithm 1 achieves acceleration and all-time constraint satisfaction at the expense of slightly more resource utilization. We compare the computation and communication load of the methods as follows. The dual subgradient method and B-DPP require each agent to exchange a single scalar variable per constraint, while Algorithm 1 requires the exchange of two scalar variables. The execution time on average per iteration of Algorithm 1 is 0.79 ms, while the other two consume 0.04 ms. This is because, for this example, Algorithm 1 solves a constrained quadratic programming problem at each iteration, while both the dual subgradient method and B-DPP solve unconstrained problems that have closed-form solutions.

6.3. Performance of the closed-loop system

In this subsection, we contrast the performance of the CBF controller under the centralized quadratic programming solver in MATLAB and the distributed solver in Algorithm 1.

The states are initialized as $z_i = z_i(0) = (2\cos(2\pi i/7) + 2, 2\sin(2\pi i/7)+1), i = 1, \ldots, N$. For the continuous-time system dynamics (15), we discretize it with a sampling period 0.01 s based on the Euler method. Regarding the implementation of

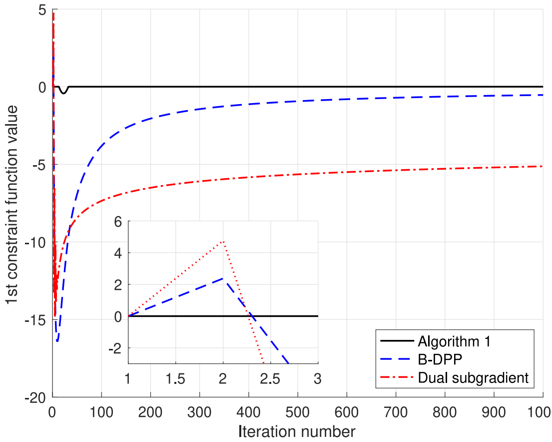


Fig. 4. Value of the first coupling constraint function $\sum_{i=1}^{4} A_i^{[1]} x_i + b_i^{[1]} \le 0$.

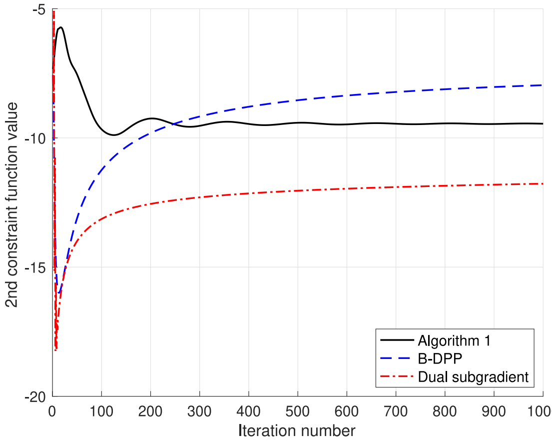


Fig. 5. Value of the second coupling constraint function $\sum_{i=4}^{7} A_i^{[2]} x_i + b_i^{[2]} \le 0$.

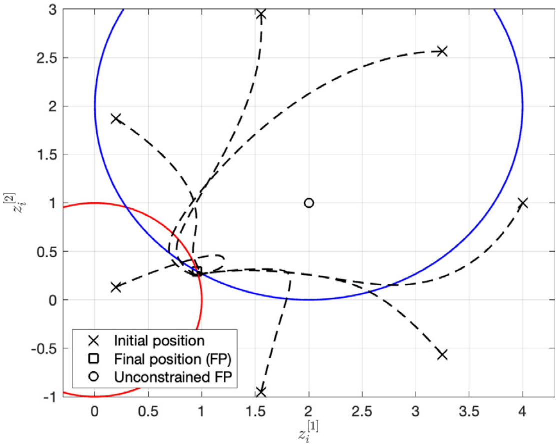


Fig. 6. Performance under the centralized CBF controller. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

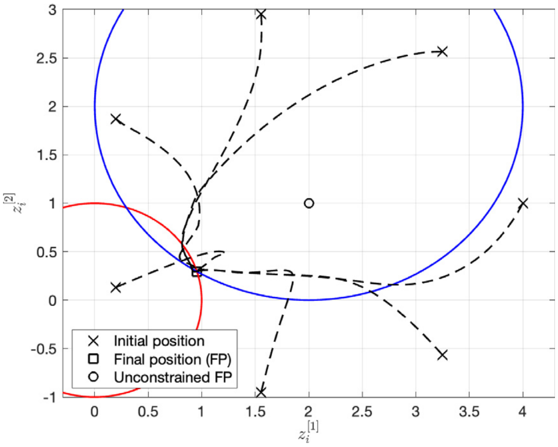


Fig. 7. Performance under the distributed CBF controller. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Algorithm 1 for the controller (18) in closed-loop, the parameter γ is manually tuned to 0.5 and kept uniform throughout. A stopping criterion is applied during each sampling period, where the algorithm terminates when the difference between two consecutive updates of $\hat{y}$ is smaller than 0.05. In the initial iteration of each sampling period, we set the auxiliary variables to be zero. We remark that Assumption 1 is universally valid, and Assumption 2 is only violated in exceptional cases when the coefficients associated with x_4 in the constraint become linearly dependent.

The comparison results are presented in Figs. 6 and 7. The blue and red curves stand for contours of functions $(z_i^{[1]})^2 + (z_i^{[2]})^2$ and $(z_i^{[1]} - 2)^2 + (z_i^{[2]} - 2)^2$ with levels 1 and 4, respectively. It can be verified from the initial values for states $z_i(0)$, $i = 1, \ldots, N$, that the constraints presented in the consensus problem, i.e., the constraints in (13), are initially violated. Nevertheless, CBF can still be used to achieve constrained consensus. Due to the violation-free feature of Algorithm 1, the CBF constraints in (16) and (17) are satisfied at all times. Under both the centralized and distributed controllers, the agents eventually meet at one of the intersections of the two curves, suggesting that constrained consensus is achieved.

7. Conclusion

This work presented a violation-free distributed optimization algorithm for networked optimization problem with coupling constraints. Convergence of the proposed algorithm was proven under suitable conditions. We studied its application to the distributed implementation of CBF controllers where constraint satisfaction is essential, and the results demonstrated the effectiveness of the approach.

This work paves the way for numerous opportunities for future research, including exploring the relaxation of the rank condition in Assumption 2-(ii) and extending the approach to address nonlinear coupling constraints. Another direction is the consideration of more practical network conditions, such as directed network, data saturation (Doostmohammadian, 2023), signed communication, and time delays (Doostmohammadian et al., 2023).

References

Alghunaim, Sulaiman A., Yuan, Kun, & Sayed, Ali H. (2019). A proximal diffusion strategy for multiagent optimization with sparse affine constraints. *IEEE Transactions on Automatic Control*, *65*(11), 4554–4567.

Ames, Aaron D, Xu, Xiangru, Grizzle, Jessy W, & Tabuada, Paulo (2016). Control barrier function based quadratic programs for safety critical systems. *IEEE Transactions on Automatic Control*, *62*(8), 3861–3876.

Beck, Amir (2017). *First-Order Methods in Optimization*. SIAM.

Bertsekas, Dimitri (2016). *Nonlinear Programming*. Athena Scientific.

Boyd, Stephen, Diaconis, Persi, & Xiao, Lin (2004). Fastest mixing Markov chain on a graph. *SIAM Review*, *46*(4), 667–689.

Boyd, Stephen, Parikh, Neal, Chu, Eric, Peleato, Borja, Eckstein, Jonathan, et al. (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*, *3*(1), 1–122.

Boyd, Stephen, & Vandenberghe, Lieven (2004). *Convex Optimization*. Cambridge University Press.

Camisa, Andrea, Farina, Francesco, Notarnicola, Ivano, & Notarstefano, Giuseppe (2021). Distributed constraint-coupled optimization via primal decomposition over random time-varying graphs. *Automatica*, *131*, Article 109739.

Cohen, Michael, Diakonikolas, Jelena, & Orecchia, Lorenzo (2018). On acceleration with noise-corrupted gradients. In *International conference on machine learning* (pp. 1019–1028). PMLR.

Doostmohammadian, Mohammadreza (2023). Distributed energy resource management: All-time resource-demand feasibility, delay-tolerance, nonlinearity, and beyond. *IEEE Control Systems Letters*.

Doostmohammadian, Mohammadreza, Aghasi, Alireza, Rikos, Apostolos I, Grammenos, Andreas, Kalyvianaki, Evangelia, Hadjicostis, Christoforos N, et al. (2022). Distributed anytime-feasible resource allocation subject to heterogeneous time-varying delays. *IEEE Open Journal of Control Systems*, *1*, 255–267.

Doostmohammadian, Mohammadreza, Aghasi, Alireza, Vrakopoulou, Maria, Rabiee, Hamid R, Khan, Usman A, & Charalambous, Themistoklis (2023). Distributed delay-tolerant strategies for equality-constraint sum-preserving resource allocation. *Systems & Control Letters*, *182*, Article 105657.

Falsone, Alessandro, Margellos, Kostas, Garatti, Simone, & Prandini, Maria (2017). Dual decomposition for multi-agent distributed optimization with coupling constraints. *Automatica*, *84*, 149–158.

Florenzano, Monique, & Le Van, Cuong (2001). *vol. 13*, *Finite dimensional convexity and optimization*. Springer Science & Business Media.

Jittorntrum, Krisorn (1978). *Sequential algorithms in nonlinear programming*. The Australian National University (Australia).

Kyparisis, Jerzy (1985). On uniqueness of Kuhn-Tucker multipliers in nonlinear programming. *Mathematical Programming*, *32*, 242–246.

Li, Xiuxian, Feng, Gang, & Xie, Lihua (2020). Distributed proximal algorithms for multiagent optimization with coupled inequality constraints. *IEEE Transactions on Automatic Control*, *66*(3), 1223–1230.

Li, Xiuxian, Xie, Lihua, & Hong, Yiguang (2019). Distributed continuous-time algorithm for a general nonsmooth monotropic optimization problem. *International Journal of Robust and Nonlinear Control*, *29*(10), 3252–3266.

Liu, Changxin, Li, Huiping, & Shi, Yang (2020). A unitary distributed subgradient method for multi-agent optimization with different coupling sources. *Automatica*, *114*, Article 108834.

Lu, Tzon-Tzer, & Shiou, Sheng-Hua (2002). Inverses of 2 × 2 block matrices. *Computers & Mathematics with Applications*, *43*(1–2), 119–129.

Mateos-Núnez, David, & Cortés, Jorge (2016). Distributed saddle-point subgradient algorithms with Laplacian averaging. *IEEE Transactions on Automatic Control*, *62*(6), 2720–2735.

Meng, Min, & Li, Xiuxian (2023). Linear last-iterate convergence for continuous games with coupled inequality constraints. In *2023 62nd IEEE conference on decision and control* (pp. 1076–1081). IEEE.

Mesbahi, Mehran, & Egerstedt, Magnus (2010). *Graph theoretic methods in multiagent networks*. Princeton University Press.

Mestres, Pol, & Cortés, Jorge (2023). Distributed and anytime algorithm for network optimization problems with separable structure. In *2023 62nd IEEE conference on decision and control* (pp. 5463–5468). IEEE.

Nedić, Angelia, & Ozdaglar, Asuman (2009). Approximate primal solutions and rate analysis for dual subgradient methods. *SIAM Journal on Optimization*, *19*(4), 1757–1780.

Notarnicola, Ivano, & Notarstefano, Giuseppe (2019). Constraint-coupled distributed optimization: a relaxation and duality approach. *IEEE Transactions on Control of Network Systems*, *7*(1), 483–492.

Shi, Wei, Ling, Qing, Wu, Gang, & Yin, Wotao (2015). Extra: An exact first-order algorithm for decentralized consensus optimization. *SIAM Journal on Optimization*, *25*(2), 944–966.

Subotić, Irina, Hauswirth, Adrian, & Dörfler, Florian (2021). Quantitative sensitivity bounds for nonlinear programming and time-varying optimization. *IEEE Transactions on Automatic Control*, *67*(6), 2829–2842.

Tan, Xiao, & Dimarogonas, Dimos V. (2021). Distributed implementation of control barrier functions for multi-agent systems. *IEEE Control Systems Letters*, *6*, 1879–1884.

Turan, Berkay, & Alizadeh, Mahnoosh (2022). Safe dual gradient method for network utility maximization problems. In *2022 IEEE 61st conference on decision and control* (pp. 6953–6959). IEEE.

Wang, Han, Papachristodoulou, Antonis, & Margellos, Kostas (2023). Distributed control design and safety verification for multi-agent systems. arXiv preprint arXiv:2303.12610.

Wang, Dandan, Zhu, Daokuan, Ou, Zichong, & Lu, Jie (2023). A distributed buffering drift-plus-penalty algorithm for coupling constrained optimization. *IEEE Control Systems Letters*.

Wu, Xuyang, Magnússon, Sindri, & Johansson, Mikael (2021). A new family of feasible methods for distributed resource allocation. In *2021 60th IEEE conference on decision and control* (pp. 3355–3360). IEEE.

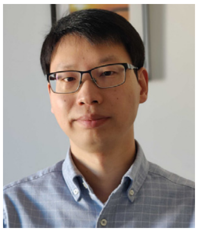
Wu, Xuyang, Magnússon, Sindri, & Johansson, Mikael (2023). Distributed safe resource allocation using barrier functions. *Automatica*, *153*, Article 111051.

Xu, Xiangru, Tabuada, Paulo, Grizzle, Jessy W, & Ames, Aaron D (2015). Robustness of control barrier functions for safety critical control. *IFAC-PapersOnLine*, *48*(27), 54–61.

Yang, Shiping, Tan, Sicong, & Xu, Jian-Xin (2013). Consensus based approach for economic dispatch problem in a smart grid. *IEEE Transactions on Power Systems*, *28*(4), 4416–4426.

Zhu, Ligeng, Liu, Zhijian, & Han, Song (2019). Deep leakage from gradients. *Advances in Neural Information Processing Systems*, *32*.

Changxin Liu received his Ph.D. in mechanical engineering from the University of Victoria, Victoria, BC, Canada, in 2021. From 2021 to 2024, he was a postdoctoral researcher at the School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, Stockholm, Sweden. In October 2024, he joined the Department of Automation at East China University of Science and Technology, Shanghai, China, where he is currently a Professor. His research interests include distributed optimization and control, cyber–physical systems security, and combinatorial optimization, with applications in the process industry.



Xiao Tan received his B.Eng. and M.Sc. degrees from Beihang University, China, in 2015 and 2018, respectively, and his Ph.D. degree in Electrical Engineering from KTH Royal Institute of Technology, Sweden, in 2023. He is currently a postdoctoral research associate at Amber Lab, California Institute of Technology, USA. His research interests lie in the areas of correct- and safe-by-construction control design for dynamical systems and optimization-based control approaches, with applications to autonomous and multi-robot systems.

Xuyang Wu received the B.S. degree in Information and Computing Science from Northwestern Polytechnical University, Xi'an, China in 2015, and the Ph.D. degree from the University of Chinese Academy of Sciences, China in 2020. He was a postdoctoral researcher in the Division of Decision and Control Systems at KTH Royal Institute of Technology, Stockholm, Sweden, from 2020 to 2023. He is currently an Associate Professor in the School of Automation and Intelligent Manufacturing at the Southern University of Science and Technology, Shenzhen, China. His research interests include distributed/parallel optimization and machine learning.

Dimos V. Dimarogonas was born in Athens, Greece, in 1978. He received the Diploma degree in electrical and computer engineering and the Ph.D. degree in mechanical engineering from the National Technical University of Athens, Athens, Greece, in 2001 and 2007, respectively. From 2007 to 2010, he was a Postdoctoral Researcher with the Department of Automatic Control, KTH Royal Institute of Technology, Stockholm, Sweden, and the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA, USA. He is currently a Professor with the Division of Decision and Control Systems, School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology. His research interests include multiagent systems, hybrid systems and control, robot navigation and manipulation, human–robot interaction, and networked control. Dr. Dimarogonas is in the Editorial Board of the Automatica and IEEE Transactions on Control of Network Systems. He was a recipient of the ERC Starting Grant in 2014, the ERC Consolidator Grant in 2019, and the Knut och Alice Wallenberg Academy Fellowship in 2015.

Karl H. Johansson is Professor with the School of Electrical Engineering and Computer Science at KTH Royal Institute of Technology in Sweden and Director of Digital Futures. He received M.Sc. degree in Electrical Engineering and Ph.D. in Automatic Control from Lund University. He has held visiting positions at UC Berkeley, Caltech, NTU, HKUST Institute of Advanced Studies, and NTNU. His research interests are in networked control systems and cyber–physical systems with applications in transportation, energy, and automation networks. He is President of the European Control Association and member of the IFAC Council, and has served on the IEEE Control Systems Society Board of Governors and the Swedish Scientific Council for Natural Sciences and Engineering Sciences. He has received several best paper awards and other distinctions from IEEE, IFAC, and ACM. He has been awarded Swedish Research Council Distinguished Professor, Wallenberg Scholar with the Knut and Alice Wallenberg Foundation, Future Research Leader Award from the Swedish Foundation for Strategic Research, the triennial IFAC Young Author Prize, and IEEE Control Systems Society Distinguished Lecturer. He is Fellow of the IEEE and the Royal Swedish Academy of Engineering Sciences.