

Accelerated Distributed Primal-Dual Dynamics Using Adaptive Synchronization

P. A. BANSODE^{1,2}, (Member, IEEE), K. C. KOSARAJU³,
S. R. WAGH¹, R. PASUMARTHY⁴, AND N. M. SINGH¹

¹Department of Electrical Engineering, Veermata Jijabai Technological Institute, Mumbai 400019, India

²Department of Instrumentation Engineering, Ramrao Adik Institute of Technology, Navi Mumbai 400706, India

³Faculty of Science and Engineering, University of Groningen, 9747 AG Groningen, The Netherlands

⁴Department of Electrical Engineering, Indian Institute of Technology Madras, Chennai 600036, India

Corresponding author: P. A. Bansode (prashant1987bansode@gmail.com)

ABSTRACT This paper proposes an adaptive primal-dual dynamics for distributed optimization in multi-agent systems. The proposed dynamics incorporates an adaptive synchronization law that reinforces the interconnection strength between the coupled agents. By strengthening the synchronization between the primal variables of the coupled agents, the given law accelerates the convergence of the proposed dynamics to the saddle-point solution. The resulting dynamics is represented as a feedback-interconnected networked system that proves to be passive. The passivity properties of the proposed dynamics are exploited along with the LaSalle's invariance principle for hybrid systems, to establish asymptotic convergence and stability of the saddle-point solution. Further, the primal dynamics is analyzed for the rate of convergence and stronger convergence bounds are established, it is proved that the primal dynamics achieve accelerated convergence under the adaptive synchronization. The robustness of the proposed dynamics is quantified using L_2 -gain analysis and the correlation between the rate of convergence and robustness of the proposed dynamics are presented. The effectiveness of the proposed dynamics is demonstrated by applying it to solve distributed least squares and distributed support vector machines problems.

INDEX TERMS Adaptive synchronization, distributed optimization, passivity, primal-dual dynamics.

I. INTRODUCTION

Distributed optimization remains a subject of substantial research over recent years. Their applications include wireless sensor networks [1]–[3], power networks [4], large scale support vector machines [5], [6] etc. An exhaustive survey of these techniques can be found in [7]. Mainly, distributed optimization techniques are categorized as either decomposition based distributed optimization (see, [8] and references therein) or consensus-based distributed optimization. The consensus based distributed optimization techniques are significantly explored lately [4]–[6], [9]–[12], which is the prime subject of this paper.

Many algorithms are proposed to solve consensus-based distributed optimization problems arising in networked systems, such as the seminal work on distributed sub-gradient methods [13], distributed primal-dual dynamical

algorithms [4], distributed gradient descent algorithms [10], [14] etc. Out of these, the distributed primal-dual dynamics based algorithms deserve special attention because of their rich systems and control theoretic properties [15]–[19] and ability to obtain simultaneously both primal as well as dual optimal solutions. The seminal work on the primal-dual dynamics or the saddle point dynamics dates back to late 1950s [20], [21]. Its application for solving optimization problems over a network first appeared in [15] with the focus on asymptotic convergence and stability of these algorithms. This framework is later extended to distributed optimization over a network of communicating nodes in [4], [22]. The primal-dual dynamics in [22] combine the decomposition and the consensus-based methods to propose proportional-integral distributed optimization for equality constrained optimization problems and achieves a globally asymptotically stable saddle-point solution. The primal-dual gradient-based algorithm proposed in [4] achieves asymptotic convergence for a consensus-based distributed optimization

The associate editor coordinating the review of this article and approving it for publication was Florin Pop.

problem with local inequality constraints and implements the algorithm for load-sharing control in power networks. The notion of asymptotic convergence and stability of the (distributed) primal-dual dynamics are well established.

From the perspectives of online optimization, the distributed algorithms must be certified based on not only the stability but also the rate of convergence. The rate of convergence of such algorithms quantifies how fast they converge to the optimal solution. Recently, the algorithms such as distributed gradient (sub-gradient) methods are widely studied with the objective of improvement in the rate of convergence, see [10], [14], [23]–[25]. However, the distributed primal-dual dynamics are not yet explored with the same objective which could limit their application to large-scale distributed optimization problems. While the existing methods on improving the rate convergence of the primal-dual dynamics rely upon increasing the convexity of the objective function by using quadratic penalty terms (augmented Lagrangian techniques) [18], their usage for solving distributed optimization problems will destroy the distributed structure of the objective function. Thus, increasing convexity by using quadratic penalties may not pose as a suitable way of improving the rate of convergence of the distributed primal-dual dynamics. The alternative route to this could be to exploit the graph-Laplacian properties of the underlying network and use adaptive coupling gains between the nodes to improve the convergence results. Addressing this issue, the present work primarily contributes to the accelerated convergence of the distributed primal-dual dynamics.

A. RELEVANT LITERATURE AND CONTRIBUTIONS

The work proposed in this paper is in the same spirit with the recent articles [4], [19]. In [4], the framework of primal-dual dynamics for network utility maximization [26] which uses Krasovskii type Lyapunov function to derive asymptotic convergence, is extended for distributed optimization with application to load sharing control in power systems. Our contribution significantly differs from [4] in the sense that the proposed dynamics is first analyzed using passivity tools of dynamical systems which then lead to its asymptotic stability when combined with the LaSalle's invariance principle of hybrid systems [27]. The advantage of passivity-based stability analysis is that the proposed dynamics can be realized as a feedback interconnection of the primal and the dual subsystems. This also facilitates to understand the interaction between the primal and the dual dynamical subsystems using port variables [19]. Thus each subsystem also enjoys L_2 stability properties of feedback connected dynamical systems. This feature later comes to the aid of robustness analysis of the proposed dynamics using L_2 -gains. The fundamental results on passivity-based stability analysis of the primal-dual dynamics are established in [19]. Our work, in a way, extends these results for the consensus-based distributed optimization problems.

The central theme of the paper, that is the adaptively coupled primal-dual dynamics is derived by integrating the

consensus protocol in the distributed primal-dual dynamics with the adaptive coupling laws motivated from the results in [29]. In [29], the adaptive synchronization technique is proved to guarantee the synchronization between the trajectories of diffusively coupled agents of a multiagent system. This technique is essentially based on modifying the coupling weights of the diffusively coupled agents as a function of the synchronization error between them. Larger values of synchronization errors result in increasing the coupling weights and vice-a-versa. In this paper, it is shown that the adaptation in the coupling weights strengthens the synchronization of the primal variables of the coupled agents. With this, the proposed work establishes results on an accelerated convergence of the proposed dynamics to the saddle point solution. While the adaptive synchronization proves to accelerate the convergence, it is shown that it affects the robustness of the proposed dynamics. By introducing exogenous inputs in the interconnected network dynamics of the primal-dual subsystems, the L_2 -gain of the proposed dynamics is analyzed and worst-case L_2 -gain is quantified in correlation with the rate of convergence. Although it is well known that the interconnected network of passive dynamical systems is inherently robust to exogenous inputs [30], our results quantify the L_2 -gain margins and establish a relation between these margins and the rate of convergence.

To summarize, the proposed work envelopes the following key points:

- 1) The proposed algorithm, designated hereafter as the adaptively synchronized distributed primal-dual dynamics (ADPDD), ensures synchronization of the network-wide primal variables to a common trajectory which is then driven to the optimal solution.
- 2) The ADPDD is posed as a negative feedback interconnection of the primal dynamical subsystem and the dual dynamical subsystems. It is proved that these subsystems remain individually passive, which subsequently, ensures the passivity and the asymptotic stability of the proposed dynamics.
- 3) The convergence rate of the ADPDD is derived and it is proved that the ADPDD has an accelerated convergence than the distributed primal-dual dynamics (DPDD).
- 4) The L_2 -gain analysis of the proposed dynamics against the exogenous disturbances is presented to show the correlation between the rate of convergence and the robustness of the proposed algorithm.

In the end, an application of the proposed algorithm to solve distributed least-squares distributed support vector machines problems along with numerical examples are discussed.

B. NOTATIONS AND PRELIMINARIES

The set \mathbb{R} (respectively $\mathbb{R}_{\geq 0}$ or $\mathbb{R}_{> 0}$) is the set of real (respectively non-negative or positive) numbers. I_n is the $n \times n$ identity matrix. $\mathbf{0}$ is a zero vector of appropriate dimensions. For a square matrix $A \in \mathbb{R}^{n \times n}$, $\text{eig}(A) = \{\lambda_1(A), \lambda_2(A), \dots, \lambda_n(A)\} \in \mathbb{R}$ represents eigenvalues of A

in an ascending order. The smallest eigenvalue of A is given by $\lambda_1(A)$. If $B \in \mathbb{R}^{m \times n}$ and $C \in \mathbb{R}^{p \times q}$ are real matrices, then $B \otimes C \in \mathbb{R}^{mp \times nq}$ is a block matrix that defines the Kronecker product of B and C .

The interaction topology in a multi-agent system is represented using an undirected graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ with $\mathcal{N} = \{1, 2, \dots, n\}$ as the set of agents and $\mathcal{E} \subseteq \mathcal{N} \times \mathcal{N}$ as the set of edges. The neighbor set of the i^{th} agent is $\mathcal{N}_i = \{q \in \mathcal{N} | (q, i) \in \mathcal{E}\}$, where $i \in \mathcal{N}$. The number of agents n is the cardinality of \mathcal{G} . Let $D \in \mathbb{R}^{n \times n}$ be the degree matrix of \mathcal{G} and $A \in \mathbb{R}^{n \times n}$ be the adjacency matrix of \mathcal{G} , with elements $a_{iq} = a_{qi} > 0, \forall (i, q) \in \mathcal{E}$, then $L = D - A$ is the Laplacian matrix of \mathcal{G} . By definition, $L \in \mathbb{R}^{n \times n}$ is a symmetric positive semidefinite matrix that encodes the connectivity of the agents and their interaction topology in \mathcal{G} .

If $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is continuously differentiable in $x \in \mathbb{R}^n$, then $\nabla_x f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is the gradient of f with respect to x . If f is twice continuously differentiable and strictly convex in x then $\mathbb{H} = \nabla_x^2 f \in \mathbb{R}_{>0}^{n \times n}$ is a symmetric positive definite matrix of second-order partial derivatives of f with respect to x .

Consider the following dynamical system

$$\dot{x} = F(x, u), \quad y = G(x, u), \quad (1)$$

where state $x \in \mathbb{R}^n$, input $u \in \mathbb{R}^m$, and output $y \in \mathbb{R}^m$, with F, G (of appropriate dimensions) sufficiently smooth and satisfying $F(0) = G(0) = 0$.

Definition 1 ([31]): The system (1) is said to be passive if there exists a positive semidefinite storage function (Lyapunov function) $V : \mathbb{R}^n \rightarrow \mathbb{R}$, continuously differentiable in x such that $\dot{V} \leq u^T y$.

In line with Definition 1, if $\dot{V} = u^T y$ strictly holds, then the system (1) is said to be lossless.

For scalars $x, y, [x]_y^+ := x$ if $y > 0$ or $x > 0$, and $[x]_y^+ := 0$ otherwise.

The remainder of the paper is mainly divided into two sections. Section II discusses the main results of the paper and Section III presents examples to validate the proposed work. Subsection II is divided as follows: Section II-A describes the consensus-based distributed optimization problem. In Subsection II-B1 the adaptive synchronization technique is elaborated. Subsection II-B2 formulates the adaptive distributed primal-dual dynamical algorithm to solve distributed optimization problem proposed in Subsection II-A. Subsections II-C and II-D present passivity and stability analysis of the proposed dynamics. In Subsection II-E the convergence bounds of the proposed algorithm are obtained and the proof for an accelerated convergence of the same is provided. Subsection II-F provides L_2 -gain analysis of the proposed dynamics and establishes a correlation between both robustness and rate of convergence of the same. Section III presents the application of the proposed dynamics to the distributed least squares and the distributed support vector machines problems. Some numerical examples of academic interests are also discussed. Section IV concludes the paper.

II. PROBLEM FORMULATION AND MAIN RESULTS

A. DISTRIBUTED OPTIMIZATION

Consider the following distributed optimization problem

$$\begin{aligned} \min_{x \in \mathbb{R}^{ln}} f(x) &= \sum_{i=1}^n f_i(x_i) \\ \text{subject to } x_{ik} &= x_{qk}, \quad \forall_{k=1}^l, \quad \forall i, q \in \mathcal{N}, \\ g_j(x_{ik}) &\leq 0, \quad \forall_{j=1}^{m_g^{ik}}, \quad \forall_{k=1}^l, \quad \forall i \in \mathcal{N}, \end{aligned} \quad (2)$$

where $x_i = [x_{i1}, \dots, x_{il}]^T \in \mathbb{R}^l$ and $x = [x_1^T, \dots, x_n^T]^T \in \mathbb{R}^{ln}$. It is assumed that the functions $f_i : \mathbb{R}^l \rightarrow \mathbb{R}$ is twice differentiable and strictly convex, and $g_j : \mathbb{R} \rightarrow \mathbb{R}$ is convex. The optimization problem (2) can be decomposed into n subproblems wherein each subproblem minimizes the cost $f_i(x_i)$ subject to the consensus constraint $x_{ik} = x_{qk}$ and inequality constraints $g_j(x_{ik}) \leq 0$. The problem (2) can not be fully decoupled into a set of n subproblems because of the consensus constraints, but it can be addressed as a network-based multiagent optimization problem using graph theory as a tool. Let an undirected and connected graph $\mathcal{G}(\mathcal{N}, \mathcal{E})$ describe the communication topology of the underlying network, where \mathcal{N} denotes the set of agents or subproblems, and \mathcal{E} denotes the set of communication links. Each agent minimizes a local cost function $f_i(x_i)$ subject to the consensus constraints $x_{ik} = x_{qk}, \forall_{k=1}^l, \forall q \in \mathcal{N}_i$ and the local inequality constraints $g_j(x_{ik}) \leq 0, \forall_{k=1}^l$. The global consensus corresponds to the optimal solution of (2), when $x_1^* = x_2^* = \dots = x_n^* = x^*$. The index m_g^{ik} is the number of inequality constraints associated with the scalar x_{ik} .

The strong duality of (2) is subject to the convexity of f and the constraint satisfaction given by the Slater's condition (see, [32]), which is as follows: Assuming that there exists an $x \in \text{relint}\mathcal{D}$ such that $g_j(x_{ik}) < 0, x_{ik} = x_{qk}, \forall_{k=1}^l, \forall q \in \mathcal{N}_i, \forall_{i=1}^n$, then x is strictly feasible, where \mathcal{D} is the domain of (2) defined as $\mathcal{D} = \text{dom}f$. The strict convexity of f implies that there exists at most one global optimal solution x^* . The Lagrangian function \mathcal{L} of the problem (2) is given by:

$$\mathcal{L}(x, \alpha, \theta) = f(x) + \alpha^T (L \otimes I_l)x + \sum_{i=1}^n \sum_{k=1}^l \sum_{j=1}^{m_g^{ik}} \theta_j^{ik} g_j(x_{ik}), \quad (3)$$

where $\alpha_{ik} \in \mathbb{R}$ is a Lagrange multiplier associated with the consensus constraint $x_{ik} = x_{qk}$ and $\theta_j^{ik} \in \mathbb{R}_+ = \{\theta_j^{ik} \in \mathbb{R} | \theta_j^{ik} \geq 0, \forall_{k=1}^l, \forall_{j=1}^{m_g^{ik}}, \forall i \in \mathcal{N}\}$ is a Lagrange multiplier associated with the inequality constraint $g_j(x_{ik}) \leq 0$. The vector notations of the respective Lagrange multipliers are $\theta \in \mathbb{R}_+^{m_g^{ik} ln}$ and $\alpha \in \mathbb{R}^{ln}$.

Remark 2: Assuming that the Slater's condition is satisfied and a strong duality holds, the saddle-point $(x^*, \alpha^*, \theta^*)$ satisfies the Karush-Kuhn-Tucker (KKT) conditions derived

the Lagrangian (3), as follows:

$$\begin{aligned} & \nabla_{x_{ik}^*} f_i(x_{ik}^*) + \sum_{q \in \mathcal{N}_i} a_{iq}(\alpha_{ik}^* - \alpha_{qk}^*) \\ & + \sum_{j=1}^{m_g^{ik}} (\theta_j^{ik})^* \nabla_{x_{ik}^*} g_j(x_{ik}^*) = 0, \quad \forall_{k=1}^l, \forall i \in \mathcal{N}, \\ & g_j(x_{ik}^*) \leq 0, (\theta_j^{ik})^* \geq 0, \quad \forall_{k=1}^l, \forall_{j=1}^{m_g^{ik}}, \forall i \in \mathcal{N}, \\ & (\theta_j^{ik})^* g_j(x_{ik}^*) = 0, \quad \forall_{k=1}^l, \forall_{j=1}^{m_g^{ik}}, \forall i \in \mathcal{N}, \\ & x_{ik}^* = x_{qk}^*, \quad \forall_{k=1}^l, \forall (i, q) \in \mathcal{N}. \end{aligned} \quad (4)$$

In order to ensure the global consensus of the states x_i , $\forall i \in \mathcal{N}$, the Lagrangian function defined in (3) is augmented with the term $x^T(L \otimes I_l)x$. The augmented Lagrangian function is defined below:

$$\tilde{\mathcal{L}}(x, \alpha, \theta) = \mathcal{L}(x, \alpha, \theta) + x^T(L \otimes I_l)x. \quad (5)$$

Remark 3: Note that augmenting the Lagrangian (3) with $x^T(L \otimes I_l)x$ does not affect its convexity-concavity properties. This owes to the fact that $x^T(L \otimes I_l)x$ is a positive semidefinite function of the primal variable x . Thus the saddle-point satisfying (4) also satisfies the following KKT conditions for the Lagrangian (5):

$$\begin{aligned} & \nabla_{x_{ik}^*} f_i(x_{ik}^*) + \sum_{q \in \mathcal{N}_i} a_{iq}(x_{ik}^* - x_{qk}^*) + \sum_{q \in \mathcal{N}_i} a_{iq}(\alpha_{ik}^* - \alpha_{qk}^*) \\ & + \sum_{j=1}^{m_g^{ik}} (\theta_j^{ik})^* \nabla_{x_{ik}^*} g_j(x_{ik}^*) = 0, \quad \forall_{k=1}^l, \forall i \in \mathcal{N}, \\ & g_j(x_{ik}^*) \leq 0, (\theta_j^{ik})^* \geq 0, \quad \forall_{k=1}^l, \forall_{j=1}^{m_g^{ik}}, \forall i \in \mathcal{N}, \\ & (\theta_j^{ik})^* g_j(x_{ik}^*) = 0, \quad \forall_{k=1}^l, \forall_{j=1}^{m_g^{ik}}, \forall i \in \mathcal{N}, \\ & x_{ik}^* = x_{qk}^*, \quad \forall_{k=1}^l, \forall (i, q) \in \mathcal{N}. \end{aligned} \quad (6)$$

Using the augmented Lagrangian (5), the primal-dual dynamics is derived as follows:

$$\begin{aligned} \dot{x}_{ik} &= -\nabla_{x_{ik}} \tilde{\mathcal{L}}(x, \alpha, \theta), \quad \dot{\alpha}_{ik} = \nabla_{\alpha_{ik}} \tilde{\mathcal{L}}(x, \alpha, \theta), \\ \dot{\theta}_j^{ik} &= [\nabla_{\theta_j^{ik}} \tilde{\mathcal{L}}(x, \alpha, \theta)]_{\theta_j^{ik}}^+, \quad \forall_{k=1}^l; \forall_{j=1}^{m_g^{ik}}; \forall i \in \mathcal{N}. \end{aligned} \quad (7)$$

With the primal-dual dynamics derived as given in (7), the following subsection develops the ADPDD.

B. ADAPTIVELY SYNCHRONIZED DISTRIBUTED PRIMAL-DUAL DYNAMICS

The following subsection presents the adaptive synchronization mechanism which is later integrated with the dynamics defined in (7) to arrive at ADPDD.

1) ADAPTIVE SYNCHRONIZATION

The adaptive synchronization mechanism is widely used in multi-agent systems to guarantee synchronization between

the agents with respect to their state variables [29], [33], which is explained subsequently.

The primal variables associated with each agent evolve according to

$$\dot{x}_{ik} = -\nabla_{x_{ik}} \tilde{\mathcal{L}}(x, \alpha, \theta) \quad (8)$$

as described in (7).

By performing gradient descent on (5), the primal dynamics (8) can be further derived as:

$$\begin{aligned} \dot{x}_{ik} &= -\nabla_{x_{ik}} f(x) - \sum_{q \in \mathcal{N}_i} a_{iq}(x_{ik} - x_{qk}) \\ & - \sum_{q \in \mathcal{N}_i} a_{iq}(\alpha_{ik} - \alpha_{qk}) - \sum_{j=1}^{m_g^{ik}} \theta_j^{ik} \nabla_{x_{ik}} g_j(x_{ik}). \end{aligned} \quad (9)$$

Let $u_{x_{ik}} \in \mathbb{R}$ corresponds to the following term in (9):

$$u_{x_{ik}} = - \sum_{q \in \mathcal{N}_i} a_{iq}(x_{ik} - x_{qk}), \quad \forall q \in \mathcal{N}_i, \quad (10)$$

where the interconnection strength or the coupling weight a_{iq} belongs to the adjacency matrix A such that

$$a_{iq} = a_{qi} = \begin{cases} \text{a positive scalar,} & \text{for } (q, i) \in \mathcal{E}, \\ 0, & \text{for } (q, i) \notin \mathcal{E}. \end{cases}$$

The equation (10) is regarded widely as the consensus protocol or the consensus law [29], [34]. Define further $u_{x_i} \in \mathbb{R}^l$, the consensus protocol (10) can be modified to accommodate $x_i \in \mathbb{R}^l$ as given below:

$$u_{x_i} = - \sum_{q \in \mathcal{N}_i} a_{iq}(x_i - x_q), \quad \forall q \in \mathcal{N}_i, \quad (11)$$

Similarly,

$$u_x = -(L \otimes I_l)x \quad (12)$$

is a compact form representation of (11).

If i and q are neighbors in \mathcal{G} with $e_{iq} = x_i - x_q$ defined as the local synchronization error, then the coupling weight can be represented as a function of e_{iq} , i.e. $\hat{a}_{iq} = h_i(e_{iq})$, where $h_i : \mathbb{R}^l \rightarrow \mathbb{R}$ monotonically increases in e_{iq} . It yields a stronger synchronization between the primal variables of the coupling agents which motivates to incorporate adaptive synchronization to address the convergence rate of the distributed primal-dual dynamics. In line with this, the following coupling weight update rule is proposed:

$$\dot{a}_{iq} = d_{iq}(e_{iq}^T e_{iq} + \dot{e}_{iq}^T \dot{e}_{iq}), \quad (13)$$

where $d_{iq} = d_{qi} > 0$ is the adaptive gain constant.

Remark 4: Represent (13) in the form $\dot{a}_{iq} = h_i(e_{iq}, \dot{e}_{iq})$, throughout the rest of the paper it is assumed that the real valued function $h_i : \mathbb{R}^{2l} \rightarrow \mathbb{R}$ is Lipschitz continuous.

The dynamics (13) incorporates two aspects of synchronization, viz. the Euclidean distance between the diffusively coupled primal variables and its derivative at a given time t . The quadratic appearance of e_{iq} and \dot{e}_{iq} in (13) ensures that it is monotonically increasing in \mathbb{R} .

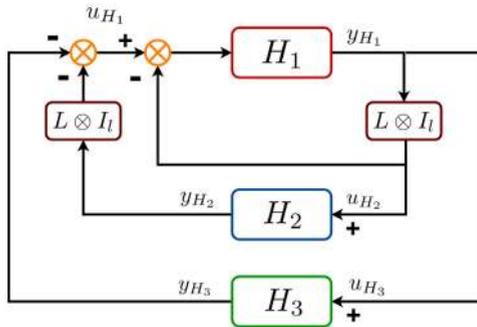


FIGURE 1. Interconnected networked dynamics of H_1 , H_2 and H_3 .

2) INTEGRATING THE ADAPTIVE COUPLING LAW (13) WITH THE PRIMAL-DUAL DYNAMICS (7)

By integrating the adaptive coupling law (13) with the PDD (7) and partitioning the resulting dynamics into three interconnected subsystems i.e., H_1 (primal partition), H_2 (consensus dual partition), and H_3 (inequality dual partition) as shown in Fig. 1, yields:

$$H_1 : \begin{cases} \dot{x} = -\nabla_x f(x) + u_x + u_{H_1}, \\ \dot{a}_{iq} = d_{iq}(e_{iq}^T e_{iq} + \dot{e}_{iq}^T \dot{e}_{iq}), \quad \forall i \in \mathcal{N}, \forall q \in \mathcal{N}_i, \\ y_{H_1} = x. \end{cases} \quad (14)$$

$$H_2 : \begin{cases} \dot{\alpha} = u_{H_2}, \\ y_{H_2} = \alpha. \end{cases} \quad (15)$$

The system H_3 represents the θ_j^{ik} dynamics in the stacked vector form with u_{H_3} and y_{H_3} as its input and output respectively, as given below:

$$H_3 : \begin{cases} \dot{\theta}_j^{ik} = [g_j(x_{ik})]_{\theta_j^{ik}}^+, \quad \forall_{k=1}^l; \forall_{j=1}^{m_g^k}; \forall i \in \mathcal{N}, \\ y_{H_3} = \sum_{j=1}^{m_g^k} \theta_j^{ik} \nabla_{x_{ik}} g_j(x_{ik}), \quad \forall_{k=1}^l; \forall_{j=1}^{m_g^k}; \\ \forall i \in \mathcal{N}, \end{cases} \quad (16)$$

where $y_{H_1}, y_{H_2}, y_{H_3} \in \mathbb{R}^n$ and $u_{H_1} = -(L \otimes I_l)y_{H_2} - y_{H_3}$, $u_{H_2} = (L \otimes I_l)y_{H_1}$, and $u_{H_3} = y_{H_1}$.

The ADPDD (14)-(16) is characterized as the feedback interconnected networked system as shown in Fig. 1. Each agent in the underlying network is diffusively coupled with its neighboring agents under the communication topology that defines the interaction between such agents on the graph $\mathcal{G}(\mathcal{N}, \mathcal{E})$. It can be noted that the network representation in Fig. 1 is independent of the graph parameters such as communication topology, number of agents, and interaction links. Irrespective of such parameters, if the graph $\mathcal{G}(\mathcal{N}, \mathcal{E})$ is connected, one can arrive at the stability results of the underlying network by only verifying its passivity properties. Towards this end, the following subsection first motivates the passivity analysis of the network shown in Fig. 1 which further leads to its closed-loop stability and robustness analysis.

C. PASSIVITY BASED STABILITY ANALYSIS OF ADPDD

This section begins with passivity analysis of the subsystems H_1, H_2, H_3 and their feedback interconnection as shown

in Fig. 1 and then moves towards the stability and robustness analysis of the said feedback interconnection. The Krasovskii type storage function is defined for each subsystem (see, [15]) which leads to a new passivity property with differentiation at both ports [35, Proposition 2]. The intuition behind this proposition is to define the Krasovskii type storage function $V(x)$ for the dynamical system defined in (1), such that $\dot{V} \leq \dot{u}^T \dot{y}$, where \dot{u} and \dot{y} are considered as port variables. This inequality shows that the map from the port input \dot{u} to the port output \dot{y} is passive. Motivated by this result, subsequently it is shown that the ADPDD is a passive system.

1) H_1 IS PASSIVE

Proposition 5: Assuming that the graph G is connected and f is strictly convex in x , if there exists $x_{eq} \in \mathbb{R}^n$ that satisfies (4), then the subsystem H_1 is passive with port variables $(\dot{y}_{H_1}, \dot{u}_{H_1})$.

Proof: Let

$$\tilde{a}_{iq} = a_{iq} - a_{iq}^* \quad (17)$$

with $a_{iq}^* > 0$ defined as follows:

$$a_{iq}^* = \begin{cases} a^* & \text{if } i \text{ and } q \text{ are neighbors in } \mathcal{G}, \\ 0 & \text{if } i \text{ and } q \text{ are not neighbors in } \mathcal{G}, \end{cases} \quad (18)$$

where a^* is a constant parameter to be selected. Consider the following storage function for the update law (13) [29].

$$W = \frac{1}{2} \sum_{i=1}^p \sum_{q=1}^p \frac{1}{d_{iq}} \tilde{a}_{iq}^2. \quad (19)$$

Differentiating (19) with respect to time yields the following:

$$\dot{W} = \sum_{i=1}^p \sum_{q=1}^p \tilde{a}_{iq} (e_{iq}^T e_{iq} + \dot{e}_{iq}^T \dot{e}_{iq}). \quad (20)$$

Acknowledging the graph symmetry and substituting for $e_{iq} = x_i - x_q$, (20) modifies to

$$\begin{aligned} \dot{W} &= \dot{x}^T (L \otimes I_l) \dot{x} - a^* \dot{x}^T (L \otimes I_l) \dot{x} \\ &\quad + x^T (L \otimes I_l) x - a^* x^T (L \otimes I_l) x, \\ &= (1 - a^*) \dot{x}^T (L \otimes I_l) \dot{x} + (1 - a^*) x^T (L \otimes I_l) x. \end{aligned} \quad (21)$$

Now, consider the following storage function for H_1 , which is a sum of Krasovskii-type storage function of x and (19):

$$V_{H_1}(x) = \frac{1}{2} \dot{x}^T \dot{x} + W. \quad (22)$$

Differentiating (22) with respect to time and using (21) yields,

$$\begin{aligned} \dot{V}_{H_1}(x) &= -\dot{x}^T \mathbb{H} \dot{x} - \dot{x}^T (L \otimes I_l) \dot{x} + (1 - a^*) \dot{x}^T (L \otimes I_l) \dot{x} \\ &\quad + (1 - a^*) x^T (L \otimes I_l) x + \dot{x}^T \dot{u}_{H_1}, \\ &= -\dot{x}^T \mathbb{H} \dot{x} - a^* \dot{x}^T (L \otimes I_l) \dot{x} + (1 - a^*) x^T (L \otimes I_l) x \\ &\quad + \dot{x}^T \dot{u}_{H_1}, \\ &\leq -(\lambda_{\min}(\mathbb{H}) + a^* \lambda_2(L \otimes I_l)) \|\dot{y}_{H_1}\|^2 \\ &\quad + (1 - a^*) \lambda_2(L \otimes I_l) \|y_{H_1}\|^2 + \dot{y}_{H_1}^T \dot{u}_{H_1}. \end{aligned} \quad (23)$$

Notice that $y_{H_1} = x$ and choosing $a^* > 1$ makes the term $(1 - a^*)\lambda_2(L \otimes I_l)\|y_{H_1}\|^2$ in (23) negative definite. Since $\lambda_{\min}(\mathbb{H}) + a^*\lambda_2(L \otimes I_l) > 0$ for a non-negative value of a^* , the inequality (23) implies that the subsystem H_1 is output strictly passive (“OSP” [30]) with respect to the port variables \dot{u}_{H_1} and \dot{y}_{H_1} . \square

Assumption 6: Throughout the rest of the paper, it is assumed that $a^* > 1$.

Remark 7: If E is defined as the incidence matrix of the undirected graph \mathcal{G} , then the Laplacian matrices $L_0 \otimes I_l$ and $L \otimes I_l$ can be defined in terms of E as:

$$L_0 \otimes I_l = EE^T \quad (24)$$

$$L \otimes I_l = EC(t)E^T \quad (25)$$

where $C(t)$ is a diagonal matrix containing the coupling weights a_{iq} . Using (17), equation (25) can be modified as given below:

$$L \otimes I_l = a^*(L_0 \otimes I_l) + E\tilde{C}(t)E^T \quad (26)$$

where $\tilde{C}(t)$ is a diagonal matrix containing \tilde{a}_{iq} . Equation (26) establishes a relation between the Laplacian matrices $L_0 \otimes I_l$ and $L \otimes I_l$ wherein $L_0 \otimes I_l$ is a Laplacian matrix whose weights are constant. This leaves the right hand side of (26) with only one variable term that is $\tilde{C}(t)$ whose coupling weights are \tilde{a}_{iq} . The Lyapunov function in (19) is defined in terms of \tilde{a}_{iq} , which leads to the output strict passivity of H_1 as stated in (23). In order to ensure that the inequality (23) holds, the input

$$\begin{aligned} u_{H_1} &= -(L \otimes I_l)y_{H_2} - y_{H_3} \\ &= -[a^*(L_0 \otimes I_l) + E\tilde{C}(t)E^T]y_{H_2} - y_{H_3} \end{aligned} \quad (27)$$

to the primal subsystem defined in (14), must be designed such that Assumption 6 is satisfied.

2) H_2 IS LOSSLESS

Proposition 8: Assuming that the graph \mathcal{G} is connected and f is strictly convex in x , if there exists $\alpha_{eq} \in \mathbb{R}^{ln}$ satisfying (4), then the subsystem H_2 is passive with port variables $(\dot{y}_{H_2}, \dot{u}_{H_2})$.

Proof: Consider a Krasovskii-type storage function for H_2 as given below:

$$V_{H_2}(\alpha) = \frac{1}{2}\dot{\alpha}^T \dot{\alpha}. \quad (28)$$

Differentiating (28) with respect to time yields,

$$\begin{aligned} \dot{V}_{H_2}(\alpha) &= \dot{\alpha}^T(L \otimes I_l)\dot{x}, \\ &= \dot{y}_{H_2}^T \dot{u}_{H_2}. \end{aligned} \quad (29)$$

Hence, the subsystem H_2 is lossless with respect to port variables \dot{u}_{H_2} and \dot{y}_{H_2} . \square

3) H_3 IS PASSIVE

In the following, H_3 is modeled as a switched dynamical system.

The dynamics in (16) becomes discontinuous when $\theta_j^{ik} = 0$ and $g_j(x_{ik}) < 0$. The value of $g_j(x_{ik})^+$ switches from $g_j(x_{ik})$

to 0. To further clarify that, (16) is reformulated below as given in Kose [20].

$$\dot{\theta}_j^{ik} = \begin{cases} g_j(x_{ik}), & \text{if } \theta_j^{ik} > 0 \text{ or } g_j(x_{ik}) > 0, \\ 0. & \end{cases} \quad (30)$$

From (30), the projection is seen to be active for the second case. Let $\mathcal{I}_i = \{1, \dots, lm_g^{ik}\}$ and $\sigma_i : [0, \infty) \rightarrow \mathcal{I}_i, \forall k = 1, \dots, l; j \in \mathcal{I}_i$ be an arbitrary switching signal. Then

$$\sigma_i(t) = \{j | \theta_j^{ik} = 0, g_j(x_{ik}) \leq 0, \forall k; \forall j \in \mathcal{I}_i\}, \quad (31)$$

represents the switching time instances when there is an active projection. Considering (31), the inequality constraint dynamics given in (16) takes the form of a switched system:

$$\dot{\theta}_j^{ik} = \begin{cases} g_j(x_{ik}), & \forall k; j \notin \sigma_i(t), \\ 0, & \forall k; j \in \sigma_i(t), \end{cases} \quad (32)$$

where $\sigma_i(t) \subset \sigma(t), \forall i=1, \dots, n$. Let V_{H_3} be the Lyapunov function associated with H_3 . It is defined as given below:

$$V_{H_3}(\theta) = \frac{1}{2} \sum_{i=1}^n \sum_{k=1}^l \sum_{j \notin \sigma_i(t)} (\dot{\theta}_j^{ik})^2. \quad (33)$$

Proposition 9: The subsystem H_3 is passive with port variables \dot{u}_{H_3} , and \dot{y}_{H_3} for each pair of switching time instances $(\tau_{\sigma_i}^+, \tau_{\sigma_i}^-)$ corresponding to (32) where $\tau_{\sigma_i}^- < \tau_{\sigma_i}^+$ such that $\sigma_i(\tau_{\sigma_i}^+) = \sigma(\tau_{\sigma_i}^-) = \sigma_i \in \mathcal{I}_i$ and $\sigma_i(\tau') \neq \sigma_i$ for $\tau_{\sigma_i}^- < \tau < \tau_{\sigma_i}^+$.

Proof: Differentiating (33) with respect to time yields,

$$\begin{aligned} \dot{V}_{H_3}(\theta) &= \sum_{i=1}^n \sum_{k=1}^l \sum_{j \notin \sigma_i(t)} \dot{\theta}_j^{ik} \ddot{\theta}_j^{ik}, \\ &= \sum_{i=1}^n \sum_{k=1}^l \sum_{j \notin \sigma_i(t)} \dot{\theta}_j^{ik} \nabla g_j(x_{ik}) \dot{x}_{ik}, \\ &\leq \sum_{i=1}^n \sum_{k=1}^l \sum_{j=1}^{m_g^{ik}} \dot{\theta}_j^{ik} \nabla g_j(x_{ik}) \dot{x}_{ik}, \\ &\leq \sum_{i=1}^n \sum_{k=1}^l \dot{x}_{ik} \sum_{j=1}^{m_g^{ik}} \dot{\theta}_j^{ik} \nabla g_j(x_{ik}). \end{aligned} \quad (34)$$

Using u_{H_3} and y_{H_3} from (16) in (34),

$$\dot{V}_{H_3}(\theta) \leq \dot{y}_{H_3}^T \dot{u}_{H_3}. \quad (35)$$

Thus,

$$\sum_{i=1}^n \sum_{k=1}^l \sum_{j \notin \sigma_i(t)} V_k(\theta_j^{ik}(\tau_{\sigma_i}^+)) - V_k(\theta_j^{ik}(\tau_{\sigma_i}^-)) \leq \int_{\tau_{\sigma_i}^-}^{\tau_{\sigma_i}^+} \dot{y}_{H_3}^T \dot{u}_{H_3} dt. \quad (36)$$

(36) ensures that the switched system (32) represents a finite family of passive systems. However, it must be ensured that the Lyapunov function V_{H_3} does not increase during the switching events. In line with this, the following two cases are considered:

- 1) It may happen for some x_{ik} in (32), that the function $g_j(x_{ik})$ goes from negative to positive through 0. This will cause the Lyapunov function to change from $V_k(\theta_j^{ik}(\tau_{\sigma_i}^-))$ to $V_k(\theta_j^{ik}(\tau_{\sigma_i}^+))$. If that happens, the Lagrangian multiplier $\theta_j^{ik} > 0$ will add a new term to $V_k(\theta_j^{ik}(\tau_{\sigma_i}))$. Since, $V_k(\theta_j^{ik}(\tau_{\sigma_i}))$ is continuous in time, (36) holds for $\tau > \tau_{\sigma_i}^-$ as well as $\tau < \tau_{\sigma_i}^+$. Hence, $V_k(\theta_j^{ik}(\tau_{\sigma_i}^+)) = V_k(\theta_j^{ik}(\tau_{\sigma_i}^-))$.
- 2) In this case the projection of k^{th} constraint for a given j becomes active, i.e., θ_j^{ik} reaches to 0 from a positive value for the k^{th} constraint of the i^{th} machine. Hence, the corresponding k^{th} term of the Lyapunov function $V_k(\theta_j^{ik})$ will disappear. In turn, the following inequality will be satisfied. $V_k(\theta_j^{ik}(\tau_{\sigma_i}^+)) < V_k(\theta_j^{ik}(\tau_{\sigma_i}^-))$.

Hence, in both the cases, the Lyapunov function $V_k(\theta_j^{ik}(\tau))$ will be non-increasing. \square

D. STABILITY ANALYSIS

Proposition 10: The interconnected network dynamics (14)-(16) is stable.

Proof: Let V be the Lyapunov function for the interconnected system represented in Fig. 1 such that

$$V = V_{H_1} + V_{H_2} + V_{H_3}. \quad (37)$$

Differentiating (37) and using (23), (29), (35) yields

$$\begin{aligned} \dot{V} &= \dot{V}_{H_1} + \dot{V}_{H_2} + \dot{V}_{H_3}, \\ &\leq -(\lambda_{\min}(\mathbb{H}) + a^* \lambda_2(L \otimes I_l)) \|\dot{y}_{H_1}\|^2 \\ &\quad + (1 - a^*) \lambda_2(L \otimes I_l) \|y_{H_1}\|^2 + \dot{y}_{H_1}^T \dot{u}_{H_1} + \dot{u}_{H_2}^T \dot{y}_{H_2} \\ &\quad + \dot{u}_{H_3}^T \dot{y}_{H_3} \\ &\leq -(\lambda_{\min}(\mathbb{H}) + a^* \lambda_2(L \otimes I_l)) \|\dot{y}_{H_1}\|^2 \\ &\quad + (1 - a^*) \lambda_2(L \otimes I_l) \|y_{H_1}\|^2 \\ &\leq 0. \end{aligned} \quad (38)$$

It verifies that the interconnected network dynamics of passive subsystems (14)-(16) is passive and thus stable.

The following result establishes the boundedness of the trajectories of (14)-(16).

Proposition 11: The trajectories of (14)-(16) are bounded for all bounded initial conditions.

Proof: To show that the trajectories of (14)-(16) are bounded, consider the following storage function:

$$\bar{V} = \frac{1}{2} \|x - x^*\|^2 + \frac{1}{2} \|\alpha - \alpha^*\|^2 + \frac{1}{2} \|\theta - \theta^*\|^2 + W. \quad (40)$$

where W is the storage function defined in (19). Differentiating (40) with respect to time yields

$$\begin{aligned} \dot{\bar{V}} &= -\nabla_x \bar{\mathcal{L}}(x, \alpha, \theta)^T (x - x^*) + \nabla_\alpha \bar{\mathcal{L}}(x, \alpha, \theta)^T (\alpha - \alpha^*) \\ &\quad + (\theta - \theta^*)^T \nabla_\theta [\bar{\mathcal{L}}(x, \alpha, \theta)]_\theta^+ + \dot{W}, \\ &\leq -\nabla_x \bar{\mathcal{L}}(x, \alpha, \theta)^T (x - x^*) + \nabla_\alpha \bar{\mathcal{L}}(x, \alpha, \theta)^T (\alpha - \alpha^*) \\ &\quad + \sum_i \sum_k \sum_j (\theta_j^{ik} - (\theta_j^{ik})^*) [g_j(x_{ik})]_{\theta_j^{ik}}^+ + \dot{W}, \\ &\leq -\nabla_x \bar{\mathcal{L}}(x, \alpha, \theta)^T (x - x^*) + \nabla_\alpha \bar{\mathcal{L}}(x, \alpha, \theta)^T (\alpha - \alpha^*) \\ &\quad + (\theta - \theta^*)^T g(x) + \dot{W}. \end{aligned} \quad (41)$$

Note that $(\theta_j^{ik} - (\theta_j^{ik})^*) g_j(x_{ik}) \geq 0, \forall j \in \sigma_i(t)$ because $g_j(x_{ik}) < 0$ and $\theta_j^{ik} = 0$ as confirmed by (32). Using first order condition of convexity-concavity of the Lagrangian function (5) and replacing \dot{W} by right-hand side of (21), (42) modifies to the following:

$$\begin{aligned} \dot{\bar{V}} &\leq -[\bar{\mathcal{L}}(x, \alpha, \theta) - \bar{\mathcal{L}}(x^*, \alpha, \theta)] + [\bar{\mathcal{L}}(x, \alpha, \theta) - \bar{\mathcal{L}}(x, \alpha^*, \theta)] \\ &\quad + [\bar{\mathcal{L}}(x, \alpha, \theta) - \bar{\mathcal{L}}(x, \alpha, \theta^*)] \\ &\quad + (1 - a^*) \dot{x}^T (L \otimes I_l) \dot{x} + (1 - a^*) x^T (L \otimes I_l) x. \end{aligned} \quad (42)$$

Since $(x^*, \alpha^*, \theta^*)$ is the saddle-point of (5), with $a^* > 1$ yields the following

$$\dot{\bar{V}} \leq 0. \quad (43)$$

which is sufficient to ensure that the trajectories of (14)-(16) are bounded. \square

In what follows, the asymptotic stability of the saddle-point solution of (14)-(16) is established. To this end, the underlying networked dynamics is represented as a hybrid system wherein H_1, H_2 are represented as continuous-time dynamical systems and H_3 is represented as a system with right-hand side discontinuity. The framework of LaSalle's invariance principle for hybrid dynamical systems (see, [27]) as stated below, provides a useful result on the convergence of (14)-(16) to the saddle point solution that satisfies (4).

Proposition 12: Consider the hybrid networked dynamics (14)-(16) and let $z = [x^T, \alpha^T, \theta^T]^T \in \mathcal{X} \subseteq \mathbb{R}^{ln(2+m_k^{ik})}$, and $\Psi \subseteq \mathcal{X}$ be compact and positively invariant. Assuming that the Lyapunov function V defined in (37) is continuously differentiable and $\dot{V} \leq 0$ along the trajectories of $z(t) \in \Psi$, every trajectory in Ψ converges to ϵ , where $\epsilon \subset \Psi$ is a maximal positive invariant set of Ψ such that

- 1) $\dot{V} = 0$ for a fixed σ .
- 2) $V_k(\theta_j^{ik}(\tau_{\sigma_i}^+)) = V_k(\theta_j^{ik}(\tau_{\sigma_i}^-))$ for a switching instance τ between $\tau_{\sigma_i}^-$ and $\tau_{\sigma_i}^+$.

\square

Proposition 12 gives the next result on the convergence of (14)-(16) to the saddle point solution that satisfies the conditions in (4).

Proposition 13: The hybrid network dynamics (14)-(16) converges to the saddle point solution x^, α^*, θ^* satisfying (4).*

Proof: From Proposition 12, for a fixed σ , $\dot{V} = 0$. Thus the primal as well as dual dynamics in (14)-(16) converge to the saddle point solution contained within the set ϵ . If $g_j(x_{ik}^*) < 0$ then $(\theta_j^{ik})^* = 0$. However, if $g_j(x_{ik}^*) > 0$, then $(\theta_j^{ik})^*$ will penalize the constraint violation by rising to a large value. Since all trajectories are bounded, it contradicts the continuity of V , thus $\dot{\theta}_j^{ik} = 0$. To this end, the solutions of (14)-(16) also satisfy the KKT conditions (4) and yield the saddle point solution $(x^*, \alpha^*, \theta^*)$. \square

Choosing $a^* > 1$ and using (12), (39) modifies to

$$\begin{aligned} \dot{\bar{V}} &\leq -(\lambda_{\min}(\mathbb{H}) + a^* \lambda_2(L \otimes I_l)) \|\dot{y}_{H_1}\|^2 \\ &\quad + (1 - a^*) \lambda_2(L \otimes I_n) \|y_{H_1}\|^2 \leq 0. \end{aligned} \quad (44)$$

Proposition 14: The saddle point solution of (5) is asymptotically stable.

Proof: The proof is straightforward from Proposition 10 and Proposition 13 and (44). \square

In the recent article, [36] the global asymptotic stability of the primal-dual dynamics is proved by using the Lyapunov function similar to that of the sum of Krasovskii-type Lyapunov function (37) and the Lyapunov function defined in (40). This result can be extended to the globally asymptotic stability of the saddle-point of (5).

Remark 15: Let $\tilde{V} : \mathbb{R}^{\ln} \times \mathbb{R}^{\ln} \times \mathbb{R}^{\ln m^k_g} \rightarrow \mathbb{R}$ denote the Lyapunov function for the ADPDD (14)-(16), given as sum of the Lyapunov functions (37) and (40) as follows:

$$\tilde{V} = V + \bar{V}. \quad (45)$$

If f is strictly convex and continuously differentiable then the trajectories of (14)-(16) converge to the saddle-point $(x^, \alpha^*, \theta^*)$ which is globally asymptotically stable. The proof of the Remark would be similar to proof the of [36, Theorem 5.1]. Hence it is omitted from here to avoid repetition.*

With the global asymptotic stability of the proposed dynamics (14)-(16) established, the subsequent section addresses its rate of convergence and its comparison with the rate of convergence with the primal-dual dynamics without adaptive weights.

E. ACCELERATED CONVERGENCE USING ADPDD

Let $\mathcal{A} \subseteq \mathbb{R}_{>0}^{|\mathcal{E}|}$ define the set of coupling weights, and $|\mathcal{E}|$ define the cardinality of the edge set \mathcal{E} . Given its definition, the Laplacian matrix $L \otimes I_l$ is a parameter varying, real and symmetric matrix, which is differentiable and uniformly continuous on \mathcal{A} . As a consequence, the following hold:

Statement 16: There exists $\Lambda > 0$ such that the spectral norm $\|L \otimes I_l\| < \Lambda, \forall a_{iq} \in \mathcal{A}, \forall q \in \mathcal{N}_i, \forall i \in \mathcal{N}$.

Statement 17: The gradient of $L \otimes I_l$ with respect to a_{iq} is bounded above by some scalar $\eta, \|\nabla L \otimes I_l\| \leq \eta, a_{iq} \in \mathcal{A}$.

Let $L_0 \otimes I_l$ be the Laplacian matrix of \mathcal{G} whose coupling weights are constant parameters, then $L_0 \otimes I_l$ results in a constant matrix.

Proposition 18: If the coupling weights evolve according to the law (13), then the following holds $\forall t > t_0$:

$$\lambda_2(L \otimes I_l) > \lambda_2(L_0 \otimes I_l). \quad (46)$$

Proof: To prove (46), it is first proved that $x^T EC(t)E^T x \geq x^T EE^T x$.

$$\begin{aligned} x^T EC(t)E^T x - x^T EE^T x &= x^T (EC(t)E^T - EE^T)x \\ &= x^T [E(C(t) - I)E^T]x. \end{aligned} \quad (47)$$

For an undirected graph $\mathcal{G}, a_{iq}(t_0) \geq 1, \forall q \in \mathcal{N}_i, \forall i \in \mathcal{N}$. Then $\forall (q, i) \in \mathcal{E}, C(t) \geq I_{|\mathcal{E}|}$. Hence,

$$C(t) - I \geq 0, \quad (48)$$

in fact, $C(t_0)$ is a diagonal matrix with the coupling weights $a_{iq}(t_0)$, thus $C(t) \geq C(t_0), \forall t > t_0$. Thus from the above

reasoning, and (48),

$$x^T EC(t)E^T \geq x^T EE^T x$$

From (24) and (25),

$$x^T L \otimes I_l x \geq x^T L_0 \otimes I_l x. \quad (49)$$

Let λ_i be the i^{th} eigenvalue in the ordered-pair of eigenvalues represented below:

$$\lambda_2 \leq \dots \leq \lambda_i \leq \dots \leq \lambda_n. \quad (50)$$

Then according to Courant-Fischer theorem [37],

$$\begin{aligned} \lambda_i(EE^T) &= \min_{x \neq 0, x \perp v_1} \frac{x^T EE^T x}{x^T x} \\ &\leq \min_{x \neq 0, x \perp v_1} \frac{x^T EC(t)E^T x}{x^T x} = \lambda_i(EC(t)E^T) \end{aligned} \quad (51)$$

where v_1 is the eigenvector (vector of all ones) corresponding to the eigenvalue $\lambda_1 = 0$. Thus for $i = 2$,

$$\begin{aligned} \lambda_2(EE^T) &\leq \lambda_2(EC(t)E^T) \\ \lambda_2(L_0 \otimes I_l) &\leq \lambda_2(L \otimes I_l) \\ \lambda_2(L \otimes I_l) &> \lambda_2(L_0 \otimes I_l), \forall t > t_0. \end{aligned} \quad (52)$$

\square

Proposition 19: If the coupling weights evolve according to (13), then the following always hold:

$$\lambda_2(L \otimes I_l) \leq \frac{\lambda_n(L \otimes I_l)}{\lambda_n(L_0 \otimes I_l)} \lambda_2(L_0 \otimes I_l). \quad (53)$$

Proof: The proof simply follows from the inequality (50). Taking the ratio of the ordered pair of eigenvalues of $L \otimes I_l$ and $L_0 \otimes I_l$, yields the following:

$$\frac{\lambda_2(L \otimes I_l)}{\lambda_2(L_0 \otimes I_l)} \leq \frac{\lambda_n(L \otimes I_l)}{\lambda_n(L_0 \otimes I_l)}, \quad (54)$$

But, for $t > t_0$, the inequality (54) strictly holds. Thus

$$\lambda_2(L \otimes I_l) \leq \frac{\lambda_n(L \otimes I_l)}{\lambda_n(L_0 \otimes I_l)} \lambda_2(L_0 \otimes I_l). \quad (55)$$

\square

In what follows, Proposition 18 and 19 are used to quantify the rate of convergence of the proposed algorithm.

QUANTIFYING THE RATE OF CONVERGENCE OF THE PROPOSED ALGORITHM (14)-(16)

By enabling a timescale separation between the evolution of trajectories of x_i and a_{iq} , the primal dynamical subsystem H_1 can be written as,

$$\dot{x} = -\nabla_x f(x) + u_x + u_{H_1}, \quad (56)$$

$$\dot{a}_{iq} = \epsilon d_{iq}(e_{iq}^T e_{iq} + \dot{e}_{iq}^T \dot{e}_{iq}), \quad \forall i \in \mathcal{N}, \forall q \in \mathcal{N}_i, \quad (57)$$

with $\epsilon \ll 1$ ensuring that the primal variable x_i evolves faster than the coupling weights a_{iq} . The primal subsystem has two control inputs u_x , to study the primal dynamics with respect to u_x in (12), let us analyze the primal subsystem H_1 when u_{H_1} is at steady state or equal to 0. With the assumption

that the coupling weight dynamics is much slower, the primal dynamics is re-written as:

$$\begin{aligned} \dot{x} &= -\nabla_x f(x) - (L \otimes I_l)x + u_{H_1}, \\ &= -F(x) + u_{H_1}, \end{aligned} \quad (58)$$

where $F(x) = \nabla_x f(x) + (L \otimes I_l)x$. Strict convexity of f can be used to prove that the primal dynamics (58) is strictly monotone for all $x \in \mathbb{R}^{ln}$ (by evaluating the Jacobian of $F(x)$, i.e. $\nabla F(x) = \mathbb{H} + L \otimes I_l \geq \mu I$, where μ is the modulus of convexity of f). Since $\mathbb{H} + L \otimes I_l > 0$ is a symmetric positive definite matrix, the Jacobian $\nabla F(x)$ is symmetric and positive definite $\forall x \in \mathbb{R}^{ln}$, it proves that $F(x)$ is strictly monotone by virtue of which the primal dynamics (58) converges to the global optimizer x^* . With f being continuously differentiable in x , the global solution is also the unique solution. Uniqueness of the primal optimizer x^* remains invariant under the adaptive coupling law (13).

The following result establishes the accelerated convergence of (58) concerning the unique optimizer x^* . Let V_{H_1} define the Lyapunov function as given below:

$$V_{H_1} = \frac{1}{2} \dot{x}^T \dot{x}. \quad (59)$$

Differentiating V_{H_1} with respect to time t ,

$$\dot{V}_{H_1} \leq -(\lambda_{\min}(\mathbb{H}) + \lambda_2(L \otimes I_l)) \dot{x}^T \dot{x} \quad (60)$$

$$\leq -\lambda_m V_{H_1} \quad (61)$$

where $\lambda_m = 2(\lambda_{\min}(\mathbb{H}) + \lambda_2(L \otimes I_l))$. Therefor,

$$V_{H_1}(x(t)) \leq V_{H_1}(x(t_0)) \exp\{-\lambda_m t\}, \quad \forall t \geq t_0. \quad (62)$$

or

$$\|x - x^*\| \leq \sqrt{2V_{H_1}(x(t_0))} \exp\{-0.5\lambda_m t\}, \quad \forall t \geq t_0. \quad (63)$$

Further, since the primal-dual dynamics has a bounded convergence with respect to the saddle point solution (see Proposition 43), using Remark 15, every initial condition $x(t_0) \in \mathbb{R}^{ln}$ approaches the optimal solution x^* faster than the usual. Thus the accelerated convergence holds globally. Considering the upper bound on $\lambda_2(L \otimes I_l)$ as given in (55), let $\lambda_2(L \otimes I_l) = \frac{\lambda_n(L \otimes I_l)}{\lambda_n(L_0 \otimes I_l)} \lambda_2(L_0 \otimes I_l)$ and $\lambda_{m_0} = 2(\lambda_{\min}(\mathbb{H}) + \lambda_2(L_0 \otimes I_l))$. Then it is seen that $\lambda_m = \lambda_{\min}(\mathbb{H}) + \frac{\lambda_n(L \otimes I_l)}{\lambda_n(L_0 \otimes I_l)} \lambda_2(L_0 \otimes I_l) \gg \lambda_{m_0}$.

Remark 20: It follows from Proposition 13 and Proposition 14 that the convergence of the primal optimizer x^ and the dual optimizers α^*, θ^* is simultaneous.*

The analysis presented below obtains a relation between the convergence rate of the proposed dynamics and its L_2 -gain.

F. ROBUSTNESS ANALYSIS OF THE NETWORK DYNAMICS CONCERNING THE EXOGENOUS INPUTS

Before proceeding with the robustness analysis of this section, it is worth noting the following remark on robustness property of the passive dynamical systems.

Remark 21: From (23), (29), and (35), it is apparent that the interconnected network dynamics comprising (14)-(16) is passive, and inherently robust to the perturbations arising in the primal and dual variables [see, Proposition 4.3.1, Remark 4.3.3 of [30]].

Remark 21 states the qualitative behavior of the proposed dynamics concerning the notion of robustness. In the following, the robustness of the proposed dynamics against exogenous inputs is quantified in terms of the L_2 -gain.

Consider without loss of generality, the new inputs to (14)-(16) as

$$\begin{aligned} H_1 : \tilde{u}_{H_1} &= u_{H_1} + \Delta u_{H_1}, \\ H_2 : \tilde{u}_{H_2} &= u_{H_2} + \Delta u_{H_2}, \\ H_3 : \tilde{u}_{H_3} &= u_{H_3} + \Delta u_{H_3}, \end{aligned} \quad (64)$$

respectively, where $\Delta u_{(\cdot)}$ corresponds to the perturbations in the input $u_{(\cdot)} \in \mathbb{R}^{ln}$. As discussed in [38], $\Delta u_{(\cdot)}$ represent additive uncertainties or disturbances such as the numerical error accumulated in the corresponding variables. In what follows, the robustness of the ADPDD is quantified using L_2 -gain analysis of dynamical systems. Let $\tilde{u} = [\tilde{u}_{H_1}^T, \tilde{u}_{H_2}^T, \tilde{u}_{H_3}^T]^T$ and $y = [y_{H_1}^T, y_{H_2}^T, y_{H_3}^T]^T$.

Proposition 22: The interconnected network dynamics (14), (15), and (16) with a_{iq} updated according to (13), remains L_2 stable with the L_2 -gain, $\gamma \leq \frac{1}{\lambda_{\min}(\mathbb{H}) + a^ \lambda_2(L \otimes I_l)}$.*

Proof: Replacing the inputs in (14)-(16) by the new ones as defined in (64), the time differential of the Lyapunov function (37) modifies to the following:

$$\begin{aligned} \dot{V} &\leq -(\lambda_{\min}(\mathbb{H}) + a^* \lambda_2(L \otimes I_l)) \|\dot{y}_{H_1}\|^2 \\ &\quad + (1 - a^*) \lambda_2(L \otimes I_n) \|y_{H_1}\|^2 + \dot{y}_{H_1}^T \dot{\tilde{u}}_{H_1} \\ &\quad + \dot{y}_{H_2}^T \dot{\tilde{u}}_{H_2} + \dot{y}_{H_3}^T \dot{\tilde{u}}_{H_3}. \end{aligned} \quad (65)$$

Acknowledging that $y_{H_1} = x$ and using (12) in (65) further yields

$$\begin{aligned} \dot{V} &\leq -(\lambda_{\min}(\mathbb{H}) + a^* \lambda_2(L \otimes I_l)) \|\dot{y}_{H_1}\|^2 \\ &\quad - (1 - a^*) y_{H_1}^T u_x + \dot{\tilde{u}}^T \dot{y}, \end{aligned} \quad (66)$$

where $\lambda_{\min}(\mathbb{H}) + a^* \lambda_2(L \otimes I_l) > 0$ since \mathbb{H} is positive definite. With $a^* > 1$, the L_2 -gain of the interconnected network dynamics, from the port input $\dot{\tilde{u}}$ to the port output \dot{y} can be calculated by setting u_x to 0. From inequality (66), the map from the input $\dot{\tilde{u}}$ to the output \dot{y} remains finite L_2 -gain stable around the saddle point x^*, α^*, θ^* , when the corresponding L_2 -gain, satisfies

$$\gamma \leq \frac{1}{(\lambda_{\min}(\mathbb{H}) + a^* \lambda_2(L \otimes I_l))}. \quad (67)$$

□

The inequality (67), clearly indicates that the L_2 gain corresponding to the adaptive distributed primal-dual dynamics reduces in margin as compared to the L_2 gain corresponding to the distributed primal-dual dynamics (without adaptive

synchronization). Using (55), one can obtain the following expression for the L_2 -gain in the worst case:

$$\underline{\gamma} = \frac{1}{(\lambda_{\min}(\mathbb{H}) + a^* \frac{\lambda_n(L \otimes I)}{\lambda_n(L_0 \otimes I)} \lambda_2(L_0 \otimes I))}. \quad (68)$$

Comparing (67) and (68), it can be found out that the L_2 -gain for the ADPDD has a reduced margin than that of the DPDD. Thus the algorithm calls for trade-off between the robustness and the accelerated convergence of the proposed dynamics. While the adaptive synchronization improves the rate of convergence of the primal-dual dynamics, it simultaneously degrades the robustness of the proposed algorithm wherein the worst-case L_2 -gain is quantified by $\underline{\gamma} (< \gamma)$ in (68).

III. APPLICATIONS AND NUMERICAL EXAMPLES

This section discusses the application of the proposed dynamics to the distributed optimization problems concerning least squares [7], [39] and support vector machines [40]. These problems are solved online over a network of wireless sensors or computing devices, in such premises the rate of convergence is a vital factor. In the following, the proposed dynamics (14)-(16) is employed to solve the distributed least squares [41] and distributed support vector machines [5], [6] problems.

A. DISTRIBUTED LEAST SQUARES

Distributed least squares problems are widely studied over recent years [12], [42], [43]. These techniques find applications in parameter estimation over wireless sensor networks [44], estimation of electro-mechanical oscillation modes of large power system networks [41], [45] etc. Each agent in the network is given a task to simultaneously and iteratively compute the same least squares solution to the linear equation $Ax = b$ where $A \in \mathbb{R}^{r_1 \times r_2}$ with $r_1 > r_2$ and $b \in \mathbb{R}^{r_1 \times 1}$.

Formally, the least squares problem is defined as given below [46]:

$$\min_x \frac{1}{2} \|Ax - b\|^2. \quad (69)$$

The objective function of the least squares problem given in (69) is not necessarily a strictly convex function, thus the existence of a global solution can not be guaranteed. In this case, the primal trajectories may synchronize and converge to a local optimizer. In order to achieve the convergence to the global optimizer, the objective function in (69) can be modified to $\frac{1}{2} \|Ax - b\|^2 + \frac{\varphi}{2} \|x\|^2$, where the quadratic term with $\varphi > 0$ ensures strictly convexity of the objective function.

1) DATA PARTITIONING

It is assumed that each agent in the network adheres to $n_r = r_1/n$ consecutive rows of A and b . For the sake of simplicity, equal partitioning of the rows of A is considered.

However, the proposed approach would hold even if the partitioning is uneven.

$$A = \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_n \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}. \quad (70)$$

where $A_i \in \mathbb{R}^{n_r \times l}$ and $A_i \in \mathbb{R}^{n_r \times 1}$.

2) DISTRIBUTED FORMULATION OF LEAST SQUARES PROBLEM

The consensus-based distributed optimization formulation of (69) would require the local estimates x_1, x_2, \dots, x_n to reach consensus on the global optimizer x^* . With data partitioning as defined above, the distributed version of the least squares problem (69) [41] is defined as

$$\begin{aligned} \min_x \quad & \sum_{i=1}^n \frac{1}{2} \|A_i x - b_i\|^2 \\ \text{subject to} \quad & x_i = x_j, \quad \forall j \in \mathcal{N}_i. \end{aligned} \quad (71)$$

3) SOLUTION TO THE DISTRIBUTED LEAST SQUARES PROBLEM (71) using ADPDD

The Lagrangian problem corresponding to (71) can be defined as

$$\mathcal{L}(x, \alpha) = \sum_{i=1}^n \|A_i x - b_i\|^2 + \alpha^T L \otimes I_l x + x^T L \otimes I_l x. \quad (72)$$

Similarly to (7), the proposed dynamics can be derived from (72) as given below:

$$\begin{aligned} H_1 : \quad & \begin{cases} \dot{x} = -A^T(Ax - b) - (L \otimes I_l)x + u_{H_1}, \\ \dot{a}_{iq} = d_{iq}(e_{iq}^T e_{iq} + \dot{e}_{iq}^T \dot{e}_{iq}), \quad \forall i \in \mathcal{N}, \forall q \in \mathcal{N}_i \\ y_{H_1} = x. \end{cases} \quad (73) \\ H_2 : \quad & \begin{cases} \dot{\alpha} = u_{H_2}, \\ y_{H_2} = \alpha. \end{cases} \quad (74) \end{aligned}$$

where $u_{H_1} = -(L \otimes I_l)y_{H_2}$ and $u_{H_2} = (L \otimes I_l)y_{H_1}$.

4) SIMULATIONS

The simulation parameters are randomly generated matrix $A \in \mathbb{R}^{100 \times 80}$ and vector $b \in \mathbb{R}^{100 \times 1}$. The network with a cyclic graph topology is assumed to comprise of 4 agents wherein each agent holds $A_i \in \mathbb{R}^{25 \times 80}$ component of A as well as the respective b_i . Each agent in the network computes $x \in \mathbb{R}^{80}$ local estimates and reaches consensus over the global solution x^* as shown in the Fig. 2. The simulations were carried out using $d_{iq} = 0.1$, the rate of convergence of (73) is compared with that of the non-adaptive version of the distributed primal-dual dynamics employed to solve the problem (71). The rate of convergence is significantly improved as shown in the Fig. 3. The global solution to (71) is also compared with the solution of the least square solver lsqin in MATLAB. The global optimizer $x_1^* = x_2^* = x_3^* = x_4^*$

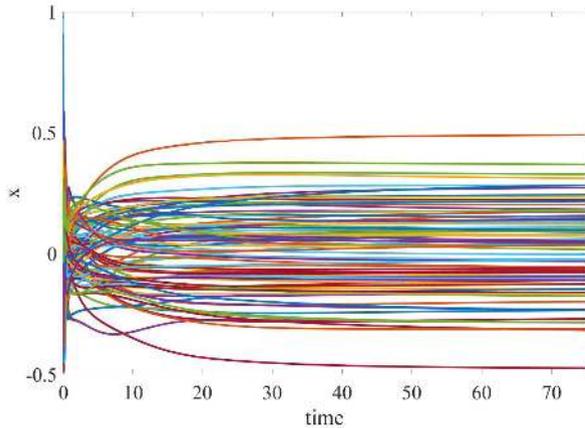


FIGURE 2. Convergence of (73) to the global solution x^* .

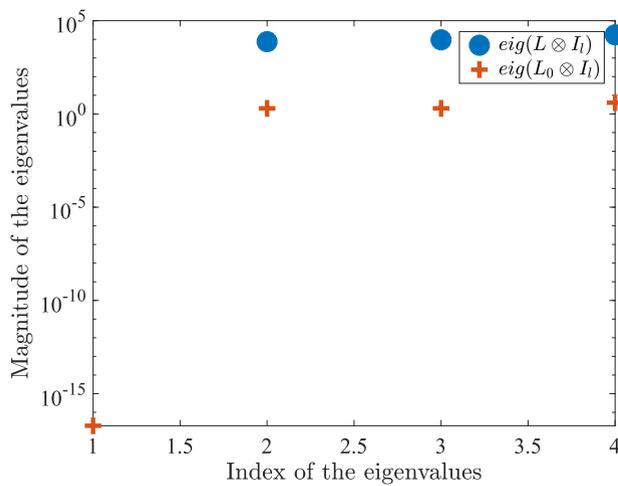


FIGURE 3. Eigenvalue comparison of $L \otimes I_l$ and $L_0 \otimes I_l$ for the problem (71).

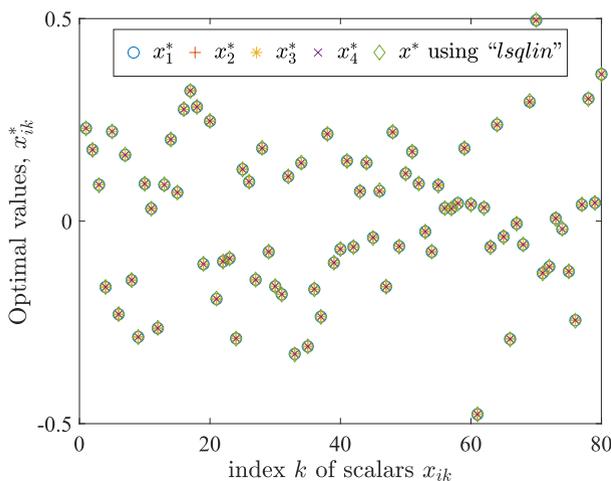


FIGURE 4. Comparison of the global optimizers of (71) with the optimal solution of the lsqlin solver.

obtained using the proposed algorithm coincides with the optimal solution x^* obtained using lsqlin as shown in the Fig. 4.

B. QUADRATIC-INEQUALITY CONSTRAINED DISTRIBUTED LEAST SQUARES

A box-constrained linear least squares problem is the one in which the upper and lower bounds on the estimated values are incorporated to handle limitations of the physical system. These methods are studied with applications to GPS positioning [47], geodesic applications [48]–[50] etc. The box-constrained least squares problem is generally defined as follows:

$$\min_x \|Ax - b\|^2, \quad \text{subject to } x_l \leq x \leq x_u, \quad (75)$$

where x_l and x_u are the upper and lower bounds of the variable x . It is known that a quadratic constraint formulation of the box constrained least square problem is an efficient approach to obtain the optimal solution of (75) [39]. The quadratic-constrained equivalent formulation of the box-constrained least square problem (76) is given as:

$$\min_x \|Ax - b\|^2, \quad \text{subject to } (x_i - \bar{x}_i)^2 \leq \rho_i^2, \quad \forall_{i=1}^n. \quad (76)$$

where \bar{x}_i is the midpoint of the interval $[x_l, x_u]$. It is computed as $\bar{x}_i = (x_l + x_u)/2$ with $\rho_i = (x_u - x_l)/2$.

A distributed framework for the quadratic-constrained least squares problem (76) can be obtained as:

$$\begin{aligned} \min_x \quad & \sum_{i=1}^n \|A_i x - b_i\|^2, \\ \text{subject to} \quad & (x_{ik} - \bar{x}_{ik})^2 \leq \rho_{ik}^2, \quad \forall_{k=1}^l, \quad \forall_{i=1}^n \\ & x_i = x_j, \quad \forall_j \in \mathcal{N}_i. \end{aligned} \quad (77)$$

The ADPDD formulation of the problem (77) is similar to that of the proposed dynamics (14)-(16). Hence, it is omitted to avoid repetition of the equations.

1) SIMULATIONS

For the sake of simplicity and readability of the simulation results, a small problem of the form (77) is taken as a proof of concept with the parameters $A \in \mathbb{R}^{20 \times 4}$ and $b \in \mathbb{R}^{20 \times 1}$. A network with a cyclic graph topology containing 4 agents is considered wherein each agent holds on to $A_i \in \mathbb{R}^{5 \times 4}$ component of the matrix A . All agents iteratively reach the global consensus of the optimizer value x^* with $d_{iq} = 2$, as shown in the Fig. 5. It can be observed that the trajectories x_1, x_2, x_3 , and x_4 synchronize to respective common trajectories at around $t \approx 0.03$ seconds. The result is also compared with the solution of lsqlin and it can be seen from the Fig. 6 that the global optimizer of (77) coincides with the solution obtained using lsqlin. The accelerated convergence of the proposed algorithm employed to solve (77) is evident from the Fig. 7.

C. DISTRIBUTED SUPPORT VECTOR MACHINES

Support vector machines (SVMs) are supervised learning-based paradigms in the machine learning domain, used for classification and regression analysis on raw data, (see [40]). For applications with a huge amount of data, there are

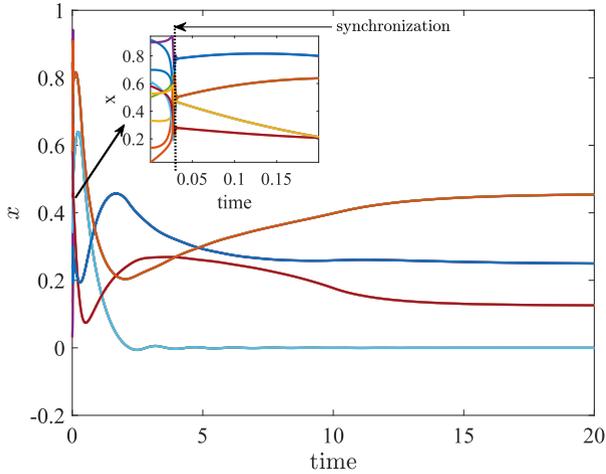


FIGURE 5. Convergence of the solutions of (77) to the global optimizer x^* .

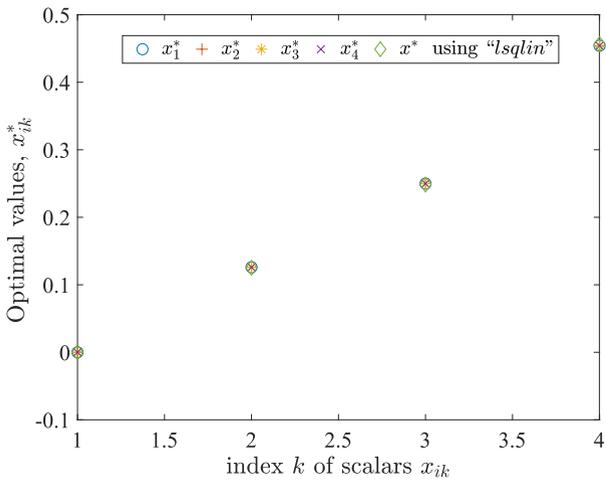


FIGURE 6. Comparison of the global optimizers of (77) with the optimal solution of the lsqlin solver.

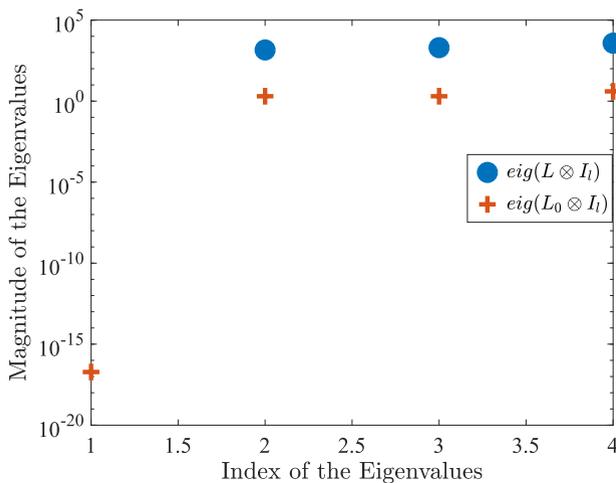


FIGURE 7. Eigenvalue comparison of $L \otimes I_l$ and $L_0 \otimes I_l$ for the problem (77).

often limitations concerning bandwidth requirement, data storage and processing capability of the computing machine, response time, etc. As it turns out, a single computing

machine is inefficient in dealing with the SVM algorithm with large datasets. Distributed versions of support vector machines are proposed as an alternative method to overcome these limitations, as discussed in [5], [6]. To enable accelerated convergence to the optimal solution, the distributed SVM problem is formulated in terms of the adaptive primal-dual dynamics. However, due to the complexity involved with simulations of large-scale SVM problems, the present work only considers the mathematical formulation and does not provide the simulation results for the same.

A problem formulation of the support vector machines for the case of non-separable data is given below:

$$\begin{aligned} \min_{w,b,\xi} \quad & \frac{1}{2} \|w\|^2 + pC \sum_{j=1}^m \xi_j \\ \text{s.t.} \quad & y_j(w^T x_j + b) \geq 1 - \xi_j, \quad \forall_{j=1}^m, \end{aligned} \quad (78)$$

where the optimization variables are weight variable w , bias variable b , and slack variable ξ . $\frac{1}{\|w\|}$ is the margin that separates positive and negative observations, $(x_j, y_j) \in S$ is a paired observation sample, respectively. $1 - \xi_j - y_j(w^T x_j + b)$ is called as a hinge loss function. C is used to trade off the sum over all slack variables ξ against the size of the margin. $p > 0$ is the scaling factor.

1) DATA PARTITIONING

It is assumed that the set of observations S is horizontally partitioned and distributed among computing nodes in $\mathcal{G}(\mathcal{N}, \mathcal{E})$ [6], where now $\mathcal{N} = \{1, \dots, n\}$ represents the computing nodes and the set of edges \mathcal{E} describes communication links between them. Assuming that the graph is connected and enabling only one-hop neighborhood communication, each node i communicates with its neighbors belonging to \mathcal{N}_i . Each node $i \in \mathcal{N}$ stores a sample set of labeled observations, denoted by $S_i = \{(x_{i1}, y_{i1}), \dots, (x_{im_i}, y_{im_i})\}$. Note that:

- 1) S_i is a set of labeled observations allocated to i^{th} computing node, $S_i \in S$, where S is a superset of the labeled observations.
- 2) $x_i \in \mathbb{R}^{m_i \times 1}$.
- 3) $y_{ij} \in \{-1, +1\}$ is a class label.

In what follows, an adaptive primal-dual dynamics based formulation of distributed support vector machines is provided.

2) ADPDD FORMULATION OF DISTRIBUTED SUPPORT VECTOR MACHINES

A distributed version of the support vector machines problem (78) is formulated as given below (see, [5]):

$$\begin{aligned} \min_{w,b,\xi} \quad & \frac{1}{2} \sum_{i=1}^n \|w_i\|^2 + pC \sum_{i=1}^n \sum_{j=1}^{m_i} \xi_{ij} \\ \text{s.t.} \quad & y_{ij}(w_i x_{ij} + b_i) \geq 1 - \xi_{ij}, \quad \xi_{ij} \geq 0, \quad \forall i \in \mathcal{N}, \quad \forall_{j=1}^{m_i} \\ & w_i = w_q, \quad b_i = b_q, \quad \forall i \in \mathcal{N}, \quad q \in \mathcal{N}_i. \end{aligned} \quad (79)$$

The objective function in (79) is a differentiable (C^2) and strongly convex in w . The decision (primal) variables are

$w, b \in \mathbb{R}^m$, where $w_i = w_q, b_i = b_q$ are the consensus constraints with q as a neighbor of i if and only if $q \in \mathcal{N}_i$. Let $h_{ij}(\xi_{ij}, w_i, b_i) = 1 - \xi_{ij} - y_{ij}(w_i x_{ij} + b_i)$.

The Lagrangian formulation of the problem (79) is given by

$$\begin{aligned} \mathcal{L}(w, b, \xi, \theta, \mu, \alpha, \beta) = & \frac{1}{2} \|w\|^2 + pC \sum_{i=1}^n \sum_{j=1}^{m_i} \xi_{ij} \\ & + \alpha^T Lw + \beta^T Lb \\ & + \sum_{i=1}^n \sum_{j=1}^{m_i} \theta_{ij} h_{ij}(\xi_{ij}, w_i, b_i) \\ & + \sum_{i=1}^n \sum_{j=1}^{m_i} \mu_{ij} \xi_{ij} + \frac{1}{2} w^T Lw + \frac{1}{2} b^T Lb, \end{aligned} \tag{80}$$

where θ_{ij}, μ_{ij} are the Lagrange multipliers associated with inequality constraints $h_{ij}(\xi_{ij}, w_i, b_i)$ and $\xi_{ij} \geq 0$, of i^{th} computing node, and α_i, β_i are the Lagrange multipliers associated with coupling constraints of i^{th} and $q^{th}, \forall q \in \mathcal{N}_i$ nodes. L is the Laplacian matrix of the undirected graph G .

Let $z = [w^T, b^T]^T$ (with $z_i = [w_i, b_i], l = 2$) then, $e_{iq} = z_i - z_q$. The interconnected network dynamics for the distributed support vector machines problem (79) is represented as follows:

$$H_1 : \begin{cases} \dot{w} = -w - Lw - L\alpha - \zeta, \\ \dot{b} = -Lb - L\beta - \eta, \\ \dot{\alpha}_{iq} = d_{iq}(e_{iq}^T e_{iq} + e_{iq}^T \dot{e}_{iq}), \quad \forall i \in \mathcal{N}, \forall q \in \mathcal{N}_i, \\ u_{H_1} = -(L \otimes I_l) y_{H_2} - y_{H_3}, \\ y_{H_1} = z. \end{cases} \tag{81}$$

The subsystem H_2 contains only consensus-dual variables, with u_{H_2} and y_{H_2} as its input and output respectively, as given below:

$$H_2 : \begin{cases} \dot{\alpha} = Lw, \\ \dot{\beta} = Lb, \\ u_{H_2} = (L \otimes I_l) y_{H_1}, \\ y_{H_2} = [\alpha^T, \beta^T]^T. \end{cases} \tag{82}$$

The subsystem H_3 contains the slack variable, and the dual variables corresponding to the inequality constraints, with u_{H_3} and y_{H_3} as its input and output respectively, as given below:

$$H_3 : \begin{cases} \dot{\theta}_{ij} = [h_{ij}(\xi_{ij}, w_i, b_i)]_{\theta_{ij}}^+ \prod_{j=1}^{m_i} \forall_{i=1}^n, \\ \dot{\mu}_{ij} = [\xi_{ij}]_{\mu_{ij}}^+ \prod_{j=1}^{m_i} \forall_{i=1}^n, \\ \dot{\xi}_{ij} = [-pC - \mu_{ij} + \theta_{ij}]_{\xi_{ij}}^+ \prod_{j=1}^{m_i} \forall_{i=1}^n, \\ u_{H_3} = y_{H_1}, \\ y_{H_3} = [\zeta^T, \eta^T]^T, \end{cases} \tag{83}$$

where $\zeta, \eta, \mu \in \mathbb{R}^n$, and $\zeta_i = \sum_{j=1}^{m_i} \theta_{ij}(-y_{ij} x_{ij})$ with $\eta_i = \sum_{j=1}^{m_i} \theta_{ij}(-y_{ij})$.

Thus, the proposed dynamics can be implemented for solving the distributed support vector machines problem (79) as

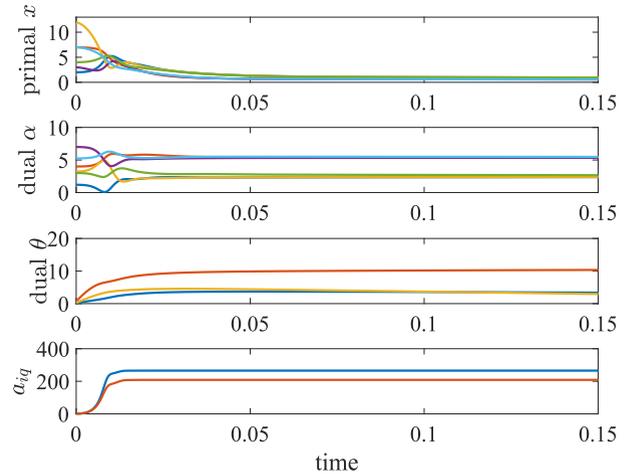


FIGURE 8. Convergence of the ADPDD (Example 1).

shown in (81)-(83). The solution of the underlying dynamics will correspond to the saddle-point solution of (80), wherein the primal solution is the optimal solution of (79).

In the following, two different formulations of (2) are considered and the results of the proposed dynamics are compared with that of the non-adaptive version of the distributed primal-dual dynamics.

D. NUMERICAL EXAMPLE 1

Consider the following distributed optimization problem consisting 3 agents having more than one variable and convex inequality constraints.

$$\begin{aligned} \min_{x \in \mathbb{R}^6} & \sum_{i=1}^3 f_i(x_i), \\ \text{subject to } & x_i = x_q, g_i(x_i) \leq 0, \quad \forall i, q \in \mathcal{N}. \end{aligned} \tag{84}$$

where the objective function associated with each agent is given below

$$f_1(x_1) = (x_{11} - x_{12})^2 + (x_{11} - 1)^2, \tag{85}$$

$$f_2(x_2) = \frac{1}{3}(x_{21} - x_{22})^2 + (x_{21} - 3)^2, \tag{86}$$

$$f_3(x_3) = \frac{1}{3}(x_{31} - x_{32})^2 + (x_{31} - 6)^2, \tag{87}$$

with the following local inequality constraints

$$g_1(x_1) = 6x_{11}^2 + 3x_{12}^2 - 11, \tag{88}$$

$$g_2(x_2) = 7x_{21}^2 + 11x_{22}^2 - 7, \tag{89}$$

$$g_3(x_3) = 2x_{31}^2 + 9x_{32}^2 - 20. \tag{90}$$

The graph connectivity is assumed to be as follows: $\mathcal{N}_1 = 1, \mathcal{N}_2 = 2$, and $\mathcal{N}_3 = 1$. The ADPDD algorithm is employed to solve the problem (84), and the corresponding trajectories are shown in Fig. 8. For $d_{iq} = 0.001$, the primal optimizers are (0.8964, 0.3538). The eigenvalues of both $L \otimes I_l$ and $L_0 \otimes I_l$ are compared as shown in Fig. 9. From Proposition 18 and

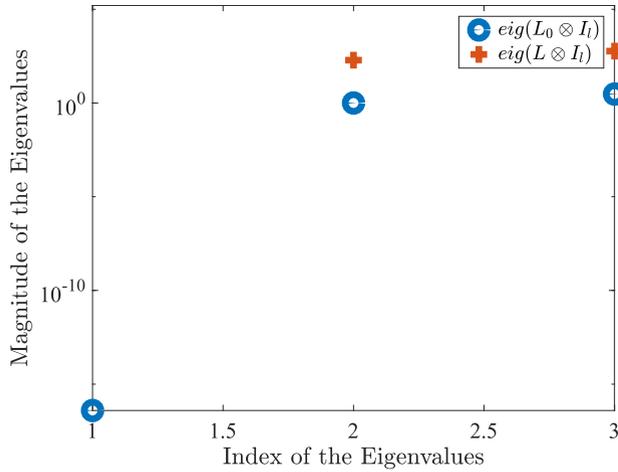


FIGURE 9. Eigenvalues of $L_0 \otimes I_l$ and $L \otimes I_l$ (Example 1).

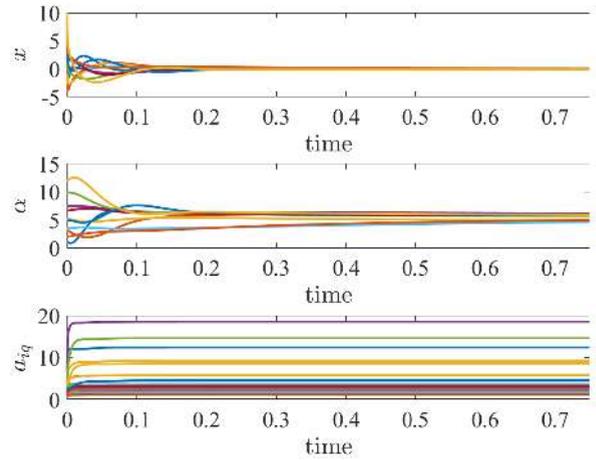


FIGURE 11. Convergence of the ADPDD ($d_{iq} = 0.001$) (Example 2).

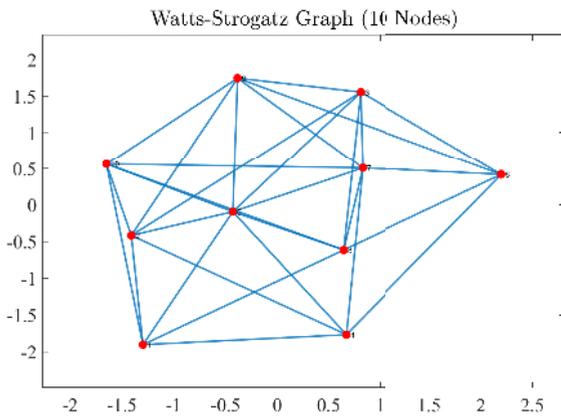


FIGURE 10. A random graph \mathcal{G} containing 10 nodes (Example 2).

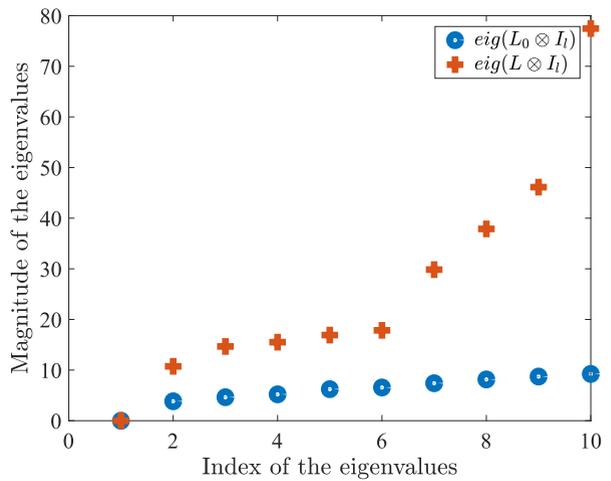


FIGURE 12. Eigenvalues of $L_0 \otimes I_l$ and $L \otimes I_l$ ($d_{iq} = 0.001$) (Example 2).

Proposition II-E, it can be seen that the adaptive synchronization has sought to increase the rate of convergence of the ADPDD.

E. NUMERICAL EXAMPLE 2

In this subsection, the local inequality constraints associated with each agent are relaxed and the following optimization problem is considered on a random graph with 10 agents as shown in Fig. 10. Note that the degree of each agent is selected randomly.

$$\begin{aligned} \min_{x \in \mathbb{R}^{10}} \quad & \sum_{i=1}^{10} f_i(x_i), \\ \text{subject to} \quad & x_i = x_q, \quad \forall i, q \in \mathcal{N}. \end{aligned} \quad (91)$$

with a randomly generated Hessian

$$\mathbb{H} = \text{diag}([136, 439, 355, 298, 302, 350, 327, 398, 353, 294]).$$

The proposed dynamics is employed to solve the optimization problem defined in (91) by first considering $d_{iq} = 0.001$ and then $d_{iq} = 0.01$. For $d_{iq} = 0.001$, Fig. 11 and Fig. 12 show the trajectories of primal-dual variables and the eigenvalues

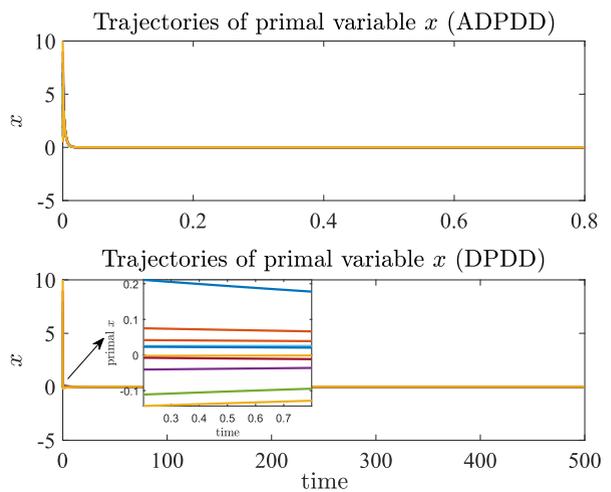


FIGURE 13. Comparison of primal variable trajectories for ADPDD and DPDD (Example 2).

of $L_0 \otimes I_l$ and $L \otimes I_l$ respectively. Similarly, Fig. 13 and Fig. 14 correspond to the trajectories of primal-dual variables and the eigenvalues of $L_0 \otimes I_l$ and $L \otimes I_l$, for the case of

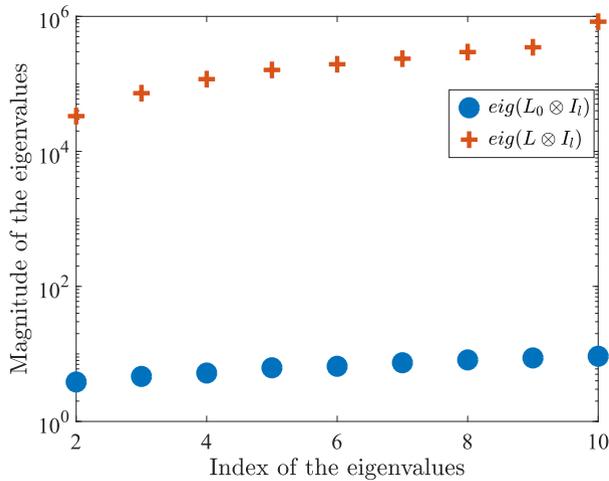


FIGURE 14. Eigenvalues of $L_0 \otimes I_l$ and $L \otimes I_l$ ($d_{iq} = 0.01$) (Example 2).

$d_{iq} = 0.01$. By comparison, it can be seen that the convergence is faster for the latter case. This owes to the difference between the resulting eigenvalues, i.e., for the case of $d_{iq} = 0.001$, the second smallest eigenvalue $\lambda_2(L \otimes I_l)$ yields to be 10.72 whereas the same for the case of $d_{iq} = 0.01$ increases to 33310. The eigenvalue results for both values of d_{iq} are shown in the Fig. 12 and the Fig. 14.

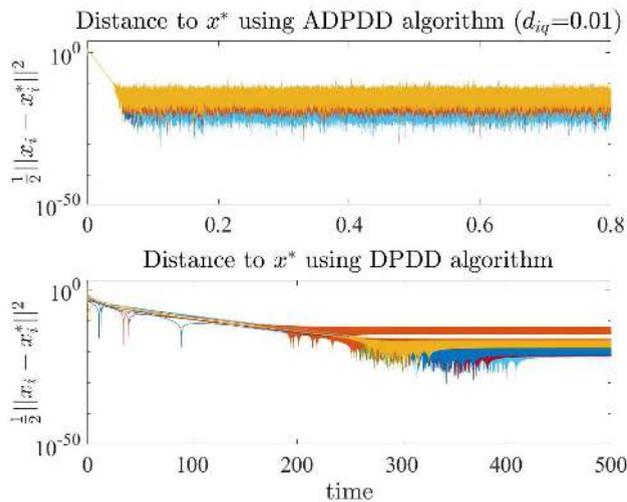


FIGURE 15. Comparison of distance to primal optimizer using ADPDD and DPDD (Example 2).

The optimal solution for the problem (91) is $x^* = \mathbf{0}$, where $\mathbf{0} \in \mathbb{R}^{10}$ is a vector containing all 0s. To show the effectiveness of the proposed algorithm, its primal variable trajectories (for $d_{iq} = 0.01$) are compared with that of the DPDD as shown in Fig. 13. The zoomed-in plot in Fig. 13 (for the time interval $[0.1, 0.8]$) depicts that the primal variable trajectories of DPDD do not synchronize to a common trajectory that will later converge to the optimal solution while for the same interval the primal variable trajectories of the ADPDD algorithm converge to the optimal solution. This result can also be verified from the distance to equilibrium plots shown in Fig. 15. A squared L_2 norm of the convergence error, i.e.

$\frac{1}{2} \|x_i - x^*\|_2^2$ is plotted against the time for both algorithms. It shows that the ADPDD converges to x^* before the sampling time $t = 0.2$ seconds whereas the DPDD converges to x^* after $t = 200$ seconds. For both algorithms, the squared L_2 norm of the convergence error is observed to be within the band 10^{-10} to 10^{-30} .

IV. CONCLUSION

In this paper, an adaptive distributed primal-dual dynamics is proposed to solve inequality and consensus constrained distributed optimization problems. The adaptive synchronization of the primal variables is brought into play by allowing the coupling weights to update according to the difference between the local trajectories (trajectories belonging to the neighboring nodes or agents) as well as the difference between the rate of change of the local trajectories respectively. It is proved that the proposed dynamics represents a network of feedback-interconnected passive dynamical systems which are asymptotically stable. Further, by allowing a time-scale separation between the adaptive coupling law and primal dynamics, stronger convergence bounds for the primal dynamic are derived, and it is proved that the adaptively coupled primal dynamics converges to the unique primal optimizer.

The performance of the proposed dynamics is quantified in terms of the induced L_2 -gain from the disturbance input to the output. The effect of adaptive synchronization on the L_2 -gain is discussed and it is established that the adaptive distributed primal-dual dynamics are comparatively less robust to the exogenous input disturbances than the distributed primal-dual dynamics. On the other hand, the analysis also revealed that to achieve accelerated convergence to the saddle-point solution, the proposed algorithm must call for a trade-off between the convergence and the robustness parameters. The future scope of the work will be directed towards developing a robust version of ADPDD algorithm with application to distributed least squares and distributed support vector machines.

REFERENCES

- [1] M. Rabbat and R. Nowak, "Distributed optimization in sensor networks," in *Proc. 3rd Int. Symp. Inf. Process. Sensor Netw.*, 2004, pp. 20–27.
- [2] B. Johansson, C. M. Carretti, and M. Johansson, "On distributed optimization using peer-to-peer communications in wireless sensor networks," in *Proc. 5th Annu. IEEE Commun. Soc. Conf. Sensor Mesh Ad Hoc Commun. Netw.*, Jun. 2008, pp. 497–505.
- [3] A. Bertrand and M. Moonen, "Consensus-based distributed total least squares estimation in ad hoc wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2320–2330, May 2011.
- [4] P. Yi, Y. Hong, and F. Liu, "Distributed gradient algorithm for constrained optimization with application to load sharing in power systems," *Syst. Control Lett.*, vol. 83, no. 9, pp. 45–52, 2015.
- [5] P. A. Forero, A. Cano, and G. B. Giannakis, "Consensus-based distributed support vector machines," *J. Mach. Learn. Res.*, vol. 11, pp. 1663–1707, May 2010.
- [6] M. Stolpe, K. Bhaduri, and K. Das, "Distributed support vector machines: An overview," in *Solving Large Scale Learning Tasks. Challenges and Algorithms*. Cham, Switzerland: Springer, 2016, pp. 109–138.
- [7] A. Nedić and J. Liu, "Distributed optimization for control," *Annu. Rev. Control Robot. Auton. Syst.*, vol. 1, pp. 77–103, May 2018.

- [8] D. P. Palomar and M. Chiang, "A tutorial on decomposition methods for network utility maximization," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1439–1451, Aug. 2006.
- [9] A. Nedić, A. Ozdaglar, and P. A. Parrilo, "Constrained consensus and optimization in multi-agent networks," *IEEE Trans. Autom. Control*, vol. 55, no. 4, pp. 922–938, Apr. 2010.
- [10] P. Lin, W. Ren, and J. A. Farrell, "Distributed continuous-time optimization: Nonuniform gradient gains, finite-time convergence, and convex constraint set," *IEEE Trans. Autom. Control*, vol. 62, no. 5, pp. 2239–2253, May 2017.
- [11] K. Sakurama and M. Miura, "Distributed constraint optimization on networked multi-agent systems," *Appl. Math. Comput.*, vol. 292, pp. 272–281, Jan. 2017.
- [12] X. Wang, J. Zhou, S. Mou, and M. J. Corless, "A distributed algorithm for least squares solutions," *IEEE Trans. Autom. Control*, to be published.
- [13] A. Nedić and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Trans. Autom. Control*, vol. 54, no. 1, pp. 48–61, Jan. 2009.
- [14] G. Qu and N. Li, "Accelerated distributed Nesterov gradient descent for smooth and strongly convex functions," in *Proc. 54th Annu. Allerton Conf. Commun. Control Comput. (Allerton)*, Sep. 2016, pp. 209–216.
- [15] D. Feijer and F. Paganini, "Stability of primal-dual gradient dynamics and applications to network optimization," *Automatica*, vol. 46, no. 12, pp. 1974–1981, 2010.
- [16] J. Wang and N. Elia, "A control perspective for centralized and distributed convex optimization," in *Proc. 50th IEEE Conf. Decis. Control Eur. Control Conf.*, Dec. 2011, pp. 3800–3805.
- [17] A. Cherukuri, E. Mallada, and J. Cortés, "Asymptotic convergence of constrained primal-dual dynamics," *Syst. Control Lett.*, vol. 87, pp. 10–15, Jan. 2016.
- [18] J. W. Simpson-Porco, "Input/output analysis of primal-dual gradient algorithms," in *Proc. 54th Annu. Allerton Conf. Commun. Control Comput. (Allerton)*, Sep. 2016, pp. 219–224.
- [19] K. C. Kosaraju, V. Chinde, R. Pasumathly, A. Kelkar, and N. M. Singh, "Stability analysis of constrained optimization dynamics via passivity techniques," *IEEE Control Syst. Lett.*, vol. 2, no. 1, pp. 91–96, Jan. 2018.
- [20] T. Kose, "Solutions of saddle value problems by differential equations," *Econometrica J. Econ. Soc.*, vol. 24, pp. 59–70, Jan. 1956.
- [21] K. J. Arrow, L. Hurwicz, and H. Uzawa, *Studies in Linear and Non-Linear Programming*. Palo Alto, CA, USA: Stanford Univ. Press, 1958.
- [22] G. Droge, H. Kawashima, and M. B. Egerstedt, "Continuous-time proportional-integral distributed optimisation for networked systems," *J. Control Decis.*, vol. 1, no. 3, pp. 191–213, 2014.
- [23] G. Qu and N. Li, "Harnessing smoothness to accelerate distributed optimization," *IEEE Trans. Control Netw. Syst.*, vol. 5, no. 3, pp. 1245–1260, Sep. 2018.
- [24] A. Nedić, A. Olshevsky, and W. Shi, "Achieving geometric convergence for distributed optimization over time-varying graphs," *SIAM J. Optim.*, vol. 27, no. 4, pp. 2597–2633, 2017.
- [25] A. Nedić, A. Olshevsky, and W. Shi, "Improved convergence rates for distributed resource allocation," in *Proc. IEEE Conf. Decis. Control (CDC)*, Dec. 2018, pp. 172–177.
- [26] D. Feijer and F. Paganini, "Krasovskii's method in the stability of network control," in *Proc. Amer. Control Conf.*, Jun. 2009, pp. 3292–3297.
- [27] J. Lygeros, K. H. Johansson, S. N. Simic, J. Zhang, and S. S. Sastry, "Dynamical properties of hybrid automata," *IEEE Trans. Autom. Control*, vol. 48, no. 1, pp. 2–17, Jan. 2003.
- [28] R. K. Brayton and J. K. Moser, "A theory of nonlinear networks. I," *Quart. Appl. Math.*, vol. 22, pp. 1–33, 1964.
- [29] S. Y. Shafi and M. Arcak, "Adaptive synchronization of diffusively coupled systems," *IEEE Trans. Control Netw. Syst.*, vol. 2, no. 2, pp. 131–141, Jun. 2015.
- [30] A. van der Schaft, *L2-Gain and Passivity Techniques in Nonlinear Control*, vol. 2. Cham, Switzerland: Springer, 2017.
- [31] H. K. Khalil, *Nonlinear Systems*, vol. 2, no. 5. Upper Saddle River, NJ, USA: Prentice-Hall, 1996, pp. 1–5.
- [32] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [33] P. De Lellis, M. di Bernardo, and F. Garofalo, "Novel decentralized adaptive strategies for the synchronization of complex networks," *Automatica*, vol. 45, no. 5, pp. 1312–1318, 2009.
- [34] M. Mesbahi and M. Egerstedt, *Graph Theoretic Methods in Multiagent Networks*, vol. 33. Princeton, NJ, USA: Princeton Univ. Press, 2010.
- [35] K. C. Kosaraju, R. Pasumathly, N. M. Singh, and A. L. Fradkov, "Control using new passivity property with differentiation at both ports," in *Proc. Indian Control Conf. (ICC)*, 2017, pp. 7–11.
- [36] A. Cherukuri, E. Mallada, S. Low, and J. Cortés, "The role of convexity in saddle-point dynamics: Lyapunov function and robustness," *IEEE Trans. Autom. Control*, vol. 63, no. 8, pp. 2449–2464, Aug. 2018.
- [37] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1990.
- [38] K. C. Kosaraju, S. Mohan, and R. Pasumathly, "On the primal-dual dynamics of support vector machines," in *Proc. 23rd Int. Symp. Math. Theory Netw. Syst.*, 2018, pp. 468–474.
- [39] J. L. Mead and R. A. Renaut, "Least squares problems with inequality constraints as quadratic constraints," *Linear Algebra Appl.*, vol. 432, no. 8, pp. 1936–1949, 2010.
- [40] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
- [41] S. Nabavi, J. Zhang, and A. Chakraborty, "Distributed optimization algorithms for wide-area oscillation monitoring in power systems using interregional PMU-PDC architectures," *IEEE Trans. Smart Grid*, vol. 6, no. 5, pp. 2529–2538, Sep. 2015.
- [42] A. H. Sayed and C. G. Lopes, "Distributed recursive least-squares strategies over adaptive networks," in *Proc. Fortieth Asilomar Conf. Signals Syst. Comput.*, Oct./Nov. 2006, pp. 233–237.
- [43] G. Mateos, I. D. Schizas, and G. B. Giannakis, "Distributed recursive least-squares for consensus-based in-network adaptive estimation," *IEEE Trans. Signal Process.*, vol. 57, no. 11, pp. 4583–4588, Nov. 2009.
- [44] S. Kar, J. M. F. Moura, and K. Ramanan, "Distributed parameter estimation in sensor networks: Nonlinear observation models and imperfect communication," *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 3575–3605, Jun. 2012.
- [45] B. Zhao, F. Lin, C. Wang, X. Zhang, M. P. Polis, and L. Y. Wang, "Supervisory control of networked timed discrete event systems and its applications to power distribution networks," *IEEE Trans. Control Netw. Syst.*, vol. 4, no. 2, pp. 146–158, Jun. 2017.
- [46] A. Björck, *Numerical Methods For Least Squares Problems*, vol. 51. Philadelphia, PA, USA: SIAM, 1996.
- [47] J. Zhu, R. Santerre, and X.-W. Chang, "A Bayesian method for linear, inequality-constrained adjustment and its application to GPS positioning," *J. Geodesy*, vol. 78, no. 9, pp. 528–534, 2005.
- [48] J. Peng, H. Zhang, S. Shong, and C. Guo, "An aggregate constraint method for inequality-constrained least squares problems," *J. Geodesy*, vol. 79, no. 12, p. 705, 2006.
- [49] X. Fang, "Weighted total least-squares with constraints: A universal formula for geodetic symmetrical transformations," *J. Geodesy*, vol. 89, no. 5, pp. 459–469, 2015.
- [50] B. Wang, J. Li, and C. Liu, "A robust weighted total least squares algorithm and its geodetic applications," *Studia Geophys. Geodaetica*, vol. 60, no. 2, pp. 177–194, Apr. 2016.



P. A. BANSODE received the M.Tech. degree in instrumentation and control engineering from the College of Engineering, Pune, India, in 2013. He is currently pursuing the Ph.D. degree in electrical engineering with the Veermata Jijabai Technological Institute, Mumbai, India. He is currently an Assistant Professor with the Faculty of Instrumentation Engineering, Ramrao Adik Institute of Technology, Mumbai. His research interests include the areas of optimization and control of networked systems, game theory, and cybersecurity.



K. C. KOSARAJU received the bachelor's degree in electronics and instrumentation from the Birla Institute of Technology and Science—Pilani, in 2010, and the master's degree in control and instrumentation and the Ph.D. degree in electrical engineering from the Indian Institute of Technology—Madras, in 2013 and 2018, respectively. He is currently a Postdoctoral with the University of Groningen, under the supervision of Professor J.M.A. Scherpen. His research interests include mainly in the areas of nonlinear control, passivity-based control and optimization theory with application to power networks, building systems, and mechanical systems.



S. R. WAGH received the Ph.D. degree in electrical engineering from The University of Western Australia, WA, Australia, in 2012. From 2015 to 2016, she was with Tufts University, Medford, MA, USA. She is currently an Assistant Professor with the Veermata Jijabai Technological Institute, Mumbai, India. Her current research interests include power system stability and control.



R. PASUMARTHY received the Ph.D. degree in systems and control from the University of Twente, The Netherlands. He is currently with the Indian Institute of Technology Madras, India, and is also associated with Robert Bosch Center for Data Sciences and Artificial Intelligence, IIT Madras. His research interests include mainly in the areas of modeling and control of complex physical systems, together with identification and control of (cloud) computing systems and data analytics for power, traffic, cloud, and brain networks.



N. M. SINGH received the Ph.D. degree in electrical engineering from IIT Bombay, Mumbai, India, in 1990. He is currently a Professor with the Veermata Jijabai Technological Institute, Mumbai. His current research interests include optimization and control of networked systems and game theory.

...