Online Distributed ADMM on Networks
Saghar Hosseini, Airlie Chapman, and Mehran Mesbahi

Abstract—This paper presents a convergence analysis on distributed Alternating Direction Method of Multipliers (ADMM) for online convex optimization problems under linear constraints. The goal is to distributively optimize a global objective function over a network of decision makers. The global objective function is composed of convex cost functions associated with each agent. The local cost functions, on the other hand, are assumed to have been broken down into two convex functions, one of which is revealed over time to the decision makers and one known a priori. In addition, the agents must achieve consensus on a global variable which is associated with a private local variable through a linear constraint. We extend an online ADMM algorithm to a distributed setting based on dual-averaging. We then propose a new performance metric for such distributed algorithms and explore the rate of convergence of the performance of the sequence of decisions generated by the algorithm as compared with the best fixed decision in hindsight. This performance metric is called the regret. A sub-linear upper bound on regret of the proposed algorithm is presented as a function of the underlying network topology and linear constraints. The online distributed ADMM algorithm is then applied to a formation acquisition problem followed by simulation results depicting its $O(\sqrt{T})$ regret bound.

Index Terms—Online Optimization; Distributed Algorithms; ADMM; Dual-averaging; Formation Algorithm

I. INTRODUCTION
A wide range of problems in engineering and information sciences can be represented as distributed convex optimization over networks, including multi-agent coordination, distributed estimation in sensor networks, decentralized tracking, and event localization. These problems often have a composite objective function subject to local linear constraints. The well known Alternating Direction Method of Multipliers (ADMM) [1] can solve this class of problems by splitting the variables as

$$\min_{x \in X, y \in Y} f(x) + \phi(y), \text{ s.t. } Ax + By = c, \quad (1)$$

where functions $f$ and $\phi$ are convex, and $\chi$ and $Y$ are convex sets.

ADMM is often considered for offline optimization problems where the cost function is known a priori. However, when the relevant decision is made, one part of the cost function might be varying with time, for example due to uncertainties in the environment and henceforth is denoted as $f_t(x)$. Further, the uncertainty in $f_t(x)$ may not be characterized by a known probability distribution. Such problem formulations fall under the class of online optimization problems [2]. Stochastic and online ADMM (O-ADMM) approaches have consequently been proposed to address this scenario which can be posed as the following optimization problem at time $T$:

$$\min_{x \in X, y \in Y} \sum_{t=1}^{T} (f_t(x) + \phi(y)), \text{ s.t. } Ax + By = c. \quad (2)$$

For this class of problems, stochastic ADMM was introduced by Ouyang et al. [3], where an identical and independent distribution for the uncertainties in function $f_t$ were considered and a convergence rate of $O(1/\sqrt{T})$ for convex functions is shown. The O-ADMM algorithms in [4], [5] are also able to provide the same convergence rate with no assumptions on the distribution of uncertainties.

Another variation of problem (1) is to consider its distributed analog,

$$\min_{x \in X, y_i \in Y} \sum_{i=1}^{n} (f_i(x) + \phi_i(y_i)), \text{ s.t. } A_ix + B_iy_i = c_i, \forall i \in [n], \quad (3)$$

and the corresponding distributed algorithm involving $n$ agents, each cooperatively solving for the global variable $x$ and their respective local variables $y_1, \ldots, y_n$. Here, the functions that compose problem (3) are distributed, specifically only agent $i$ has access to the functions $f_i$ and $\phi_i$. A motivating application for this problem is forest firefighting using a group of autonomous vehicles. As shown in Figure 1, the agents have to reach consensus on the location of wildfire represented by a global variable. Moreover, the location of
each vehicle is represented by a local variable and the group must coordinate with each other through a constraint. The detailed formulation of this application can be found in §V.

A special case of this problem formulation is the ADMM form of the consensus problem [6] where agreement is required on each agent’s local variable $y_i$, formally

$$\min_{x \in X, y_1, \ldots, y_n \in Y} \sum_{i=1}^n \phi_i(y_i), \text{ s.t. } x = y_i, \text{ for all } i \in [n]. \quad (4)$$

The constraint set can also be reformulated to represent the coupling among agents imposed by the underlying network topology. Wei and Ozdaglar [7] have proposed a stochastic asynchronous edge based ADMM algorithm to solve this problem. We observe that the objective in (4) is a local formulation of the global objective (3) and hence is more readily achieved distributively when, during each iteration of the algorithm, the constraint set might be violated. In other words, each agent is penalized by its local cost rather than the global cost, facilitating a rapid $O(1/T)$ convergence rate reported in [7]. This problem has also been examined in the context of gradient based distributed optimization [8], [9], [10], where, under the global objective (3) and local objective (4), the rate of convergence of $O(1/\sqrt{T})$ and $O(1/T)$ can be achieved, respectively.

In addition, Mota et al. [11] have studied the consensus problem in connected bipartite graphs based on distributed ADMM. Using quantitative analysis, they have shown that this algorithm requires less communication between agents compared with other algorithms to achieve a given accuracy. Deng et al. [12] have proposed a proximal Jacobian ADMM, which is suitable for parallel computation. However, this method requires an all-to-all communication in each iteration.

In this work, by fusing the online and distributed ADMM problems we examine the online distributed ADMM (OD-ADMM) at time $T$:

$$\min_{y_1, \ldots, y_n \in Y} \sum_{t=1}^T \left[ \sum_{i=1}^n \phi_i(y_i) \right], \quad (5)$$

s.t. $A_i x + B_i y_i = c_i$ for all $i \in [n].$

Inspired by the O-ADMM algorithm and our previous work [10], a single loop OD-ADMM based on dual-averaging is proposed in this paper. We consider an online convex optimization over a network of agents where each agent has two sets of variables. Similar to distributed ADMM (D-ADMM), the agents in our setup are required to reach agreement on the global variable. However, each agent keeps a private set of variables satisfying a local linear constraint which presents a relation between the global and local variables. The cost function associated with the global variable is revealed to the decision maker after committing to a decision, while the cost function associated with the local variables is known a priori. The rate of convergence of this online algorithm is shown to be $O(1/\sqrt{T})$ and is described as its regret. Regret is a metric that measures the difference between the incurred cost of the algorithm and the cost of the best fixed decision in hindsight. In the literature, the average regret of a “good” online algorithm is sub-linear with time.

The outline of the paper is as follows. In §II, the notation and a brief background on graphs and regret are presented. The optimization problem formulation and the measure of performance are stated in §III followed by the description of the OD-ADMM algorithm and the corresponding regret analysis in §IV. Then in §V, a distributed formation acquisition problem is solved based on the proposed algorithm, and simulation results are presented to support the analysis. Finally, concluding remarks are provided in §VI.

II. BACKGROUND AND PRELIMINARIES

In this section, we review basic concepts from graph theory and online algorithms, as well as the relevant assumptions for our analysis.

The notation $v_i$ or $[v]_i$ denotes the $i$th element of a column vector $v \in \mathbb{R}^p$. A unit vector $e_i$ denotes the column vector which contains all zero entries except $e_i$. The vector of all ones will be denoted by 1. For a matrix $M \in \mathbb{R}^{p \times q}$, $[M]_{ij}$ denotes the element in its $i$th row and $j$th column. A doubly stochastic matrix $P$ is a non-negative matrix with $\sum_{i=1}^n P_{ij} = \sum_{j=1}^n P_{ij} = 1$. For any positive integer $n$, the set $\{1, \ldots, n\}$ is denoted by $[n]$. The 2-norm, 1-norm and infinity norm are denoted by $\|\cdot\|_2$, $\|\cdot\|_1$, and $\|\cdot\|_\infty$, respectively, and the dual 2-norm of a vector $u$ is defined as $\|u\|_* = \sup \{\langle u, v \rangle \mid \|u\| = 1\}$. We denote the largest, second largest, and smallest singular values of $Q \in \mathbb{R}^{n \times n}$ by $\sigma_1(Q)$, $\sigma_2(Q)$ and $\sigma_n(Q)$, respectively. A function $f : \chi \to \mathbb{R}$ is called $L$-Lipschitz continuous if there exists a positive constant $L$ for which

$$|f(u) - f(v)| \leq L\|u - v\| \text{ for all } u, v \in \chi. \quad (6)$$

A. Graphs

A graph is an abstraction for representing the interactions among decision-makers, e.g., sensors and mobile robots. A weighted graph $G = (V, E, W)$ is defined by a node set $V$ where the number of nodes in the graph is $|V| = n$. Nodes represent the decision-makers in the network, and the edge set $E$ represents the agents’ interactions, that is, an agent $i$ communicates with agent $j$ if there is an edge from $i$ to $j$, i.e., $(i, j) \in E$. In addition, a weight $w_{ji} \in W$ can be associated with every edge $(i, j) \in E$ through the function $W : E \to \mathbb{R}$. The neighborhood set of node $i$ is defined as $N(i) = \{j \in V \mid (i, j) \in E\}$. One way to represent $G$ is through the adjacency matrix $A(G)$ where $[A(G)]_{ij} = w_{ji}$ for $(i, j) \in E$ and $[A(G)]_{ji} = 0$, otherwise. For a graph $G$, $d_i$ is the weighted in-degree of $i$ defined as $d_i = \sum_{(j, i) \in E} w_{ij}$. Another matrix representation of $G$ is the weighted graph Laplacian defined as $L(G) = \Delta(G) - A(G)$, where $\Delta(G)$ is the diagonal matrix of node in-degree’s $d_i$. If there exists a directed path between every pair of distinct vertices, the graph $G$ is referred to as strongly connected.
B. Regret

In online optimization, an online algorithm generates a sequence of decisions \( \langle x_t \rangle \). At iteration \( t \), the convex cost function \( l_t \) is unknown before committing to \( x_t \). The feedback available to the algorithm is the loss \( l_t(x_t) \) and its gradient. We can capture the performance of online algorithms by a standard measure called regret. Regret measures how competitive the algorithm is with respect to the best fixed solution. This best fixed decision, denoted as \( x^* \), is chosen with the benefit of hindsight. Formally, the regret is defined as the difference between the incurred cost \( l_t(x_t) \) and the cost of the best fixed decision \( l_t(x^*) \) after \( T \) iterations, i.e.,

\[
R_T = \sum_{t=1}^{T} (l_t(x_t) - l_t(x^*)).
\]

An online algorithm performs well if its regret grows sub-linearly with the number of iterations, i.e.,

\[
\lim_{T \to \infty} R_T/T = 0.
\]

This implies that the average loss of the algorithm tends to the average loss of the best fixed strategy in hindsight independent of the uncertainties associated with global cost only revealed to each decision-maker after it commits to a decision. We refer to [13], [14], [15], [16] for further discussions on online algorithms and their regret analysis.

III. PROBLEM STATEMENT

In this section, we consider a large scale network of agents cooperatively optimizing a global objective function. Let the communication geometry amongst the \( n \) decision-makers, or agents, be denoted by a graph \( G = (V, E) \). Each node \( i \in V \) is an agent which communicates with its neighbor \( j \in N(i) \) through edge \( (i,j) \in E \). An equivalent online distributed convex optimization problem to (5) is as follows

\[
\min_{x \in \chi, \; y_1, \ldots, y_n \in Y} \sum_{t=1}^{T} F_t(x,y) := \sum_{t=1}^{T} (f_t(x) + \frac{1}{n} \sum_{i=1}^{n} \phi_i(y_i))
\]

subject to

\[
r_t(x,y_i) := A_t x + B_t y_i - c_t = 0 \quad \forall i \in [n],
\]

where \( f_t(x) = \frac{1}{n} \sum_{i=1}^{n} f_{i,t}(x) \), and \( f_{i,t}(x) : \mathbb{R}^{d_x} \to \mathbb{R} \) and \( \phi_i(y_i) : \mathbb{R}^{d_y} \to \mathbb{R} \) are closed, proper, and convex cost functions associated with agent \( i \in [n] \) as shown in Figure 2. We assume an optimal solution to (9) exists over the convex sets \( \chi \subseteq \mathbb{R}^{d_x} \) and \( Y \subseteq \mathbb{R}^{d_y} \) and the optimum value of the cost function is \( P^* \).

The local decision made by agent \( i \) is represented by the optimization variables \( x_i \in \chi \) and \( y_i \in Y \). The matrices in the local linear constraints are denoted as \( A_i \in \mathbb{R}^{m_i \times d_x} \), \( B_i \in \mathbb{R}^{m_i \times d_y} \), and \( c_i \in \mathbb{R}^{m_i} \) at node \( i \in [n] \). We assume that \( B_i^T \) is left invertible, i.e., \( \sigma_{m_i}(B_iB_i^T) \) is non-zero, for all \( i \in [n] \). Let \( f_{i,t} \) and \( \phi_i \) be Lipschitz continuous with Lipschitz constant \( L_f \) and \( L_\phi \), respectively, which is defined as

\[
|f_{i,t}(u) - f_{i,t}(v)| \leq L_f \|u - v\| \quad \forall u, v \in \chi,
\]

and

\[
|\phi_i(u') - \phi_i(v')| \leq L_\phi \|u' - v'\| \quad \forall u', v' \in Y.
\]

For the online framework, each decision maker \( i \) at time \( t \) selects a global variable \( x_{i,t} \in \chi \) and local variable \( y_{i,t} \in Y \), based on local information. The cost \( f_{i,t}(x_{i,t}) \) is revealed to the agent \( i \) after its local decision \( x_{i,t} \) has been executed at time \( t \).

A. Regret for Constrained Optimization

In this section we propose a measure for evaluating the performance of OD-ADMM based on variational inequalities. This measure is inspired by the convergence analysis of Douglas-Rachford ADMM first presented in [17].

Consider the Lagrangian \( L_T \) for the constrained optimization problem (8) as

\[
L_T(x,y,\lambda) = \sum_{t=1}^{T} [f_t(x) + \frac{1}{n} \sum_{i=1}^{n} \phi_i(y_i) + \langle \lambda_t, r_t(x,y_i) \rangle],
\]

where \( x \in \chi \) and \( y_i \in Y \), as well as assuming \( \lambda \in \mathbb{Z} \), for all \( i \in [n] \). Then, the Lagrange dual function is defined as

\[
D(\lambda) = \inf_{x \in \chi, y_i \in Y} L_T(x,y,\lambda),
\]

which implies that \( D(\lambda) \) is concave and yields a lower bound on the optimal value of (8) [6]. Therefore, the goal is to maximize \( D(\lambda) \) with variable \( \lambda \in \mathbb{Z}^n \):

\[
D^* = \max_{\lambda \in \mathbb{Z}^n} D(\lambda) = \max_{\lambda \in \mathbb{Z}^n} L_T(x^*,y^*,\lambda).
\]

Based on saddle point theorem, if for all \( w = (x,y_1,\ldots,y_n,\lambda_1,\ldots,\lambda_n) \in \chi \times Y^n \times \mathbb{Z}^n = \Omega \), there exists \( \lambda^* \in \mathbb{Z}^n \) such that

\[
L_T(x^*,y^*,\lambda) \leq L_T(x,y^*,\lambda^*) \leq L_T(x,y,\lambda^*) \leq L_T(x,y,\lambda^*)
\]

i.e. \( w^* = (x^*,y^*_1,\ldots,y^*_n,\lambda^*_1,\ldots,\lambda^*_n) \in \Omega \) is a saddle point for \( L_T \), then \( (x^*,y^*_1,\ldots,y^*_n) \in \chi \times Y^n \) solves the primal problem (8)-(9) and \( \lambda^* \in \mathbb{Z}^n \) solves the dual problem (12) [18]. We assume the dual optimal solution is bounded, \( \|\lambda^*_i\| \leq D_\lambda \) for all \( i \in [n] \). A consequence of inequality (13)
is that \( \bar{\omega} = (\bar{x}, \bar{y}, \bar{\lambda}) \in \Omega \) approximately solves the primal problem with accuracy \( \epsilon_T \geq 0 \) if it satisfies
\[
0 \leq \mathcal{L}_T(\bar{x}, \bar{y}, \bar{\lambda}) - \mathcal{L}_T(x^*, y^*, \lambda^*) \leq \epsilon_T^P, \\
0 \leq \mathcal{L}_T(\bar{x}, \bar{y}, \bar{\lambda}) - \mathcal{P}^* \leq \epsilon_T^P.
\]
(14)

Based on (11), the inequality (14) can also be referred as dual feasibility. In addition, \( \bar{\omega} = (\bar{x}, \bar{y}, \bar{\lambda}) \in \Omega \) approximately solves the dual problem with accuracy \( \epsilon_T^D \geq 0 \) if
\[
0 \leq \mathcal{L}_T(x^*, y^*, \lambda^*) - \mathcal{L}_T(x^*, y^*, \bar{\lambda}) \leq \epsilon_T^D, \\
0 \leq D^* - D(\bar{\lambda}) \leq \epsilon_T^D.
\]
(15)

which represents the dual sub-optimality. The conditions in (14) and (15) can be represented as
\[
\sum_{t=1}^T f_t^\Delta(\bar{w}, w^*) + \frac{1}{n} \sum_{i=1}^n \phi_i^\Delta(\bar{w}, w^*) + H_t^\Delta(\bar{w}, w^*)) \leq \epsilon_T,
\]
where
\[
f_t^\Delta(w, w^*) = f_t(x) - f_t(x^*)
\]
\[
\phi_i^\Delta(w, w^*) = \phi_i(y) - \phi_i(y^*)
\]
\[
H_t^\Delta(w, w^*) = \mathcal{H}_t^\Delta(w, w^*) + \mathcal{H}_t^\Delta(w, w^*)
\]
\[
\mathcal{H}_t^\Delta(w, w^*) = \langle x - x^*, A_T^T \lambda_t^* \rangle + \langle \lambda_i - \lambda_i^* - r_i(x^*, y_t^*) \rangle
\]
\[
\mathcal{H}_t^\Delta(w, w^*) = \langle y_i - y_i^*, B_t^T \lambda_t^* \rangle,
\]
and \( \epsilon_T = \epsilon_T^P + \epsilon_T^D \geq 0 \).

Analogous to the regret definition O-ADMM algorithm [19], we can consider a sequence of decisions \( \{w_t\} \), where \( w_t \in \Omega \) instead of a fixed decision \( \bar{w} \). Consequently, \( \{w_t\} \) approximately solves (8) and (9) with accuracy \( \epsilon_T \) if
\[
\sum_{t=1}^T f_t^\Delta(w_t, w^*) + \frac{1}{n} \sum_{i=1}^n \phi_i^\Delta(w_t, w^*) + H_t^\Delta(w_t, w^*) \leq \epsilon_T,
\]
for the optimal solution \( w^* \in \Omega \), referred to as fixed case solutions to distinguish them from the time-varying online solutions \( \{w_t\} \). Moreover, the mapping \( H_t^\Delta(w, w^*) \) can be expressed as
\[
H_t^\Delta(w, w^*) = (i w - i w^*, H_t(w^*))
\]
where \( i w = [x \ y_i \ \lambda_i]^T \) and
\[
H_i(w) = \begin{bmatrix}
0 & 0 & A_T^T \\
0 & 0 & -B_t^T \\
-A_i & -B_i & 0
\end{bmatrix}
\]
\[
i w + \begin{bmatrix}
0 \\
0 \\
c_i
\end{bmatrix}.
\]

Since, the mapping \( H_i(w) \) is affine in \( i w \) and is defined through a skew symmetric matrix, it is monotone, and consequently [20]
\[
\langle i w - i w^*, H_i(w) - H_i(w^*) \rangle \geq 0
\]
\[
\langle i w - i w^*, H_i(w) \rangle \geq \langle i w - i w^*, H_i(w^*) \rangle.
\]
(17)

Therefore, the inequality
\[
\sum_{t=1}^T f_t^\Delta(w_t, w^*) + \frac{1}{n} \sum_{i=1}^n \phi_i^\Delta(w_t, w^*) + \langle i w_t - i w^*, H_t(w_t) \rangle \leq \epsilon_T
\]
is a sufficient condition for (16).

Finally, motivated by the inclusion of regularization terms in the augmented Lagrangian method [18], the term on the left hand side of (18) is supplemented with terms of the form \( \|r_t(x_{i,t}, y_{i,t})\|^2 \), where \( \rho > 0 \), to promote agents satisfying the local primal feasibility constraints. In our setting, the sequence \( \{w_t\} \) is constructed from the distributed algorithm adopted by the agents, specifically \( w_t = (x_{i,t}, y_t, \lambda_{i,t}) \in \Omega \) at time \( t \), where \( x_t = \frac{1}{n} \sum_{i=1}^n x_{i,t} \), \( y_t = (y_{1,t}, \cdots, y_{n,t}) \) and \( \lambda_{i,t} = (\lambda_{1,t+1}, \cdots, \lambda_{n,t+1}) \). The regret is thus defined as
\[
R_T = \max_{w \in \Omega} \sum_{t=1}^T f_t^\Delta(w_t, w^*) + \frac{1}{n} \sum_{i=1}^n \phi_i^\Delta(w_t, w^*) + \langle i w - i w^*, H_i(w_t) \rangle + \frac{\rho}{2} \|r_t(x_{i,t}, y_{i,t})\|^2.
\]
(19)

Based on (18) we say that \( \{w_t\} \) approximately solves (8) and (9) with accuracy \( \epsilon_T \) if it satisfies \( R_T \leq \epsilon_T \). Therefore, if the regret is sub-linear with time, the online algorithm will perform as well as the best fixed case decision provided with the full sequence of cost functions a priori. In addition, this property of the regret ensures that the local linear constraints will be satisfied as \( T \to \infty \).

IV. MAIN RESULT

The main contribution of this paper is adapting O-ADMM [5] and Nesterov’s dual averaging algorithm [21] to provide a distributed decision-making processes for the online optimization problem discussed in §III. The proposed algorithm updates variables \( (x_t, y_t, z_t, \lambda_t) \) for each agent \( i \in [n] \) by alternately minimizing the Lagrangian and augmented Lagrangian. In addition, the Lagrangian is linearized based on the dual averaging update which is a gradient descent method followed by a projection step onto the constraint set \( \chi \). Specifically, \( z_{t+1} = z_t + g_t \), where \( g_t = \nabla L_t(x_t) \) followed by
\[
x_{t+1} = \prod_{\chi}(z_{t+1}, \alpha_t);
\]
(20)
in this case, the parameter \( \alpha_t \) is a non-increasing sequence of positive functions and \( \prod_{\chi}(\cdot) \) is the projection operator onto \( \chi \) defined as
\[
\prod_{\chi}(z_{t+1}, \alpha_t) = \arg \min_{x \in \chi} \left\{ \langle z_{t+1}, x \rangle + \frac{1}{\alpha_t} \psi(x) \right\}.
\]
(21)

Note that proximal function \( \psi(x) : \chi \to \mathbb{R} \) is continuously differentiable and strongly convex. It acts as a regularizer to avoid oscillations in the projection step.

Finally, the proposed online algorithm minimizes the augmented Lagrangian over \( y \) as
\[
y_{t+1} = \arg \min_{y \in Y} \left\{ \mathcal{L}_t(x_{t+1}, y, \lambda_{t+1}) + \frac{\rho}{2} \|r(x_{t+1}, y)\|^2 \right\},
\]
and update the dual variable \( \lambda \) as
\[
\lambda_{t+1} = \lambda_{t+1} + \rho(A x_{t+1} + B y_{t+1} - c).
\]
1Note that the index for the dual variable is one time step ahead of the primal variables.
The distributed algorithm can be considered as an approximate ADMM by an agent $i$ via a convex combination of information provided by its neighbors $N(i)$. Specifically, the global update step (20) can be reformulated with a distributed dual averaging method. The underlying communication network can be represented compactly as a doubly stochastic matrix $P \in \mathbb{R}^{n \times n}$ which preserves the zero structure of the Laplacian matrix $L(G)$. It is clear that for agents to have access to all sub-gradients of $g_{i,t} = \nabla L_{i,t}(x_{i,t})$ there must be a information flow amongst the agents. Therefore, the graph $G$ must be strongly connected to attain this requirement. A method to construct a doubly stochastic matrix $P$ of the required form from the Laplacian of the network is provided in Proposition 3.

The Online Distributed ADMM (OD-ADMM) is presented in Algorithm 1. The projection operator $\prod_\chi^\psi(\cdot)$ used in this algorithm is defined as in (21).

**Algorithm 1: Online Distributed ADMM (OD-ADMM)**

1. Input: $\rho > 0$, $\{\alpha_i\}_{i=1}^n$
2. Initialize $z_{i,1} = \lambda_{i,1} = 0$ and $x_{i,1} = 0$, $y_{i,1} = 0$ for $\forall i = 1, ..., n$
3. for $t = 1$ to $T$
   4. foreach Agent $i$
      5. Compute subgradient $g_{i,t}(x_{i,t}) = \partial f_{i,t}(x_{i,t})$
      6. $\lambda_{i,t+1} =:\lambda_{i,t} + \rho(A_{i,x_{i,t}} + B_{i,y_{i,t}} - c_i)$
      7. $z_{i,t+1} = \sum_{j \in N(i)} P_{j,i} z_{j,t} + g_{i,t} + A_{i,t}^T \lambda_{i,t+1}$
      8. $x_{i,t+1} = \prod_\chi^\psi(z_{i,t+1}, \alpha_i)$
      9. $r_i(x_{i,t+1}, y) = A_{i,x_{i,t+1}} + B_{i,y} - c_i$
      10. $y_{i,t+1} = \arg\min_{y \in Y} \{r_i(x_{i,t+1}, y) + \frac{\rho}{2} \| r_i(x_{i,t+1}, y) \|^2 \}$
   11. end
   12. end

Before presenting the convergence rate of the proposed OD-ADMM algorithm we provide a few preliminary remarks and definitions. The sequences of average dual sub-gradient $\langle z_t \rangle$ and average sub-gradient $\langle g_t \rangle$, are defined as

$$z_t = \frac{1}{n} \sum_{i=1}^n z_{i,t}, \quad g_t = \frac{1}{n} \sum_{i=1}^n g_{i,t}. \quad (22)$$

Thus, the following update rule is introduced similar to the standard dual averaging algorithm

$$z_{t+1} = z_t + g_t + \sum_{i=1}^n A_{i,t}^T \lambda_{i,t+1}. \quad (23)$$

where the primal variable is

$$\theta_{t+1} = \prod_\chi^\psi(z_{t+1}, \alpha_t). \quad (24)$$

The regret analysis can now be presented as follows where the intermediate results required for its proof are relegated to the Appendix, namely Lemmas 4, 5 and 6.

**Theorem 1.** Given the sequence $\langle w_t \rangle$ generated by Algorithm 1 with $\psi(x^*) \leq \psi^2$ and $\alpha(t) = k/\sqrt{t}$, we have

$$R_T \leq J_1 + J_2 k \sqrt{T}, \quad (25)$$

where

$$J_1 = \frac{3D_\lambda}{2\rho n} \sum_{i=1}^n \frac{\zeta_i}{\sigma_i(A_i)},$$

$$J_2 = \zeta(4QL_f + 2Q\zeta_{\max} + 10L_f + 3\zeta) + 2\zeta + 5L_f^2 + 2QL_f^2 + \frac{\psi^2}{k^2},$$

where

$$\zeta = \frac{1}{n} \sum_{i=1}^n \zeta_i, \quad \zeta = \frac{1}{n} \sum_{i=1}^n \zeta_i^2, \quad \zeta_{\max} = \max_i \zeta_i,$$

$$\zeta_i = \sqrt{m_{\lambda}(A_i)} / \sigma_m(B_i^T).$$

Proof: Since $f_t$ is convex,

$f_t^\Delta(w_t, w^*) \leq \frac{1}{n} \sum_{i=1}^n (f_i(x_{t,i}) - f_i(x^*)) + \frac{1}{n} \sum_{i=1}^n f_i^\Delta(w_{i,t}, w^*),$ where $w_{i,t} = (x_{i,t}, y_{i,t}, \lambda_{i,t+1})$ and so

$$R_T \leq \frac{1}{n} \sum_{i,t} f_i^\Delta(w_{i,t}, w^*) + \phi_i^\Delta(w_t, w^*) + \frac{\rho}{2} \| r_i(x_{i,t}, y_{i,t}) \|^2.$$ As $f_t$ is $L$-Lipschitz and convex, we have

$$f_t^\Delta(w_{i,t}, w^*) = f_t(x_{i,t}) - f_t(\theta_t) + f_t(\theta_t) - f_t(x^*) \leq L \| x_{i,t} - \theta_t \| + \langle g_t, \theta_t - x^* \rangle. \quad (26)$$

In order to further bound the first term in (26) we can use Lemma 4 in the Appendix which implies that

$$\| x_{i,t} - \theta_t \| \leq \alpha_t \| z_t - x_{i,t} \|_* \quad (27)$$

From the integral test with $\alpha_t = k/\sqrt{t}$ it follows that

$$\sum_{t=1}^T \alpha_t \leq 2k \sqrt{T}. \quad (28)$$

Moreover, applying Lemma 6, from (27) - (28) it follows that

$$\sum_{t=1}^T \| x_{i,t} - \theta_t \| \leq 2k \sqrt{T}(Q(L_f + \zeta_{\max}) + 2L + \zeta + \zeta_i). \quad (29)$$

Applying Lemma 5 to the second term in (26) with (23) and (24) then

$$\sum_{t=1}^T \langle g_t, \theta_t - x^* \rangle \leq \sum_{t=1}^T \| \theta_t \|_2 + \frac{1}{n} \sum_{i=1}^n A_{i,t}^T \lambda_{i,t+2} \| x^* \| + \frac{1}{n} \sum_{i=1}^n \| A_{i,t+1} (x^* - \theta_t) \| + \frac{1}{\alpha_T} \psi(x^*). \quad (30)$$

Note that $\frac{1}{\sqrt{T}}$ is a non increasing positive function and the integral test leads to $\sum_{t=1}^T \| x_{i,t} - \theta_t \| \leq 2 \sqrt{T} - 1.$
The first term on the right hand side of (30) is bounded\(^3\) as
\[
\sum_{t=1}^{T} \frac{\alpha_t}{2} \| g_{t+1} + \frac{1}{n} \sum_{i=1}^{n} A_i^T \lambda_{i,t+2} \|^2 \leq \\
(\max_t \| \tilde{g}_{t+1} \|_2 + \frac{1}{n} \sum_{i=1}^{n} \sigma_i(A_i) \max_t \| \lambda_{i,t} \|)^2 \sum_{t=1}^{T} \frac{\alpha_t}{2}. \tag{31}
\]

We now proceed to bounding the individual terms in (31) separately. By optimality of line 11 in Algorithm 1 and applying line 7, we have
\[
\nabla_y \phi_t(y_{i,t}) = -B_i^T (\lambda_{i,t} + \rho r_i(x_{i,t}, y_{i,t})) = -B_i^T \lambda_{i,t+1},
\]
for all \( i \in [n] \) and \( t \in [T] \). Moreover, since \( \| \nabla_y \phi_t(y_{i,t}) \| \leq L_{\phi}, \) hence \( \| B_i^T \lambda_{i,t+1} \| \leq L_{\phi} \). Thus, \( \| \lambda_{i,t} \| \) is bounded as
\[
\| \lambda_{i,t} \| \leq \| (B_i^T B_i)^{-1} B_i \|_F \| B_i^T \lambda_{i,t+1} \| \\
\leq L_{\phi} \left( \sum_{j=1}^{m} \frac{1}{\sigma_j^2(B_i^T)} \right)^{1/2} \leq \sqrt{m} L_{\phi} \| \lambda_{t+1} \|, \tag{32}
\]
which implies that \( \| A_i^T \lambda_{i,t} \| \leq \sqrt{m} L_{\phi} \| \lambda_{t+1} \| / \sigma_{m_i}(B_i^T). \) Combining bounds (28), (32) and \( \| g_{t+1} \|_2 \leq L_f, \) into (31) then
\[
\sum_{t=1}^{T} \frac{\alpha_t}{2} \| g_{t+1} + \frac{1}{n} \sum_{i=1}^{n} A_i^T \lambda_{i,t+2} \|^2 \leq (L_f + \zeta)^2 k \sqrt{T}. \tag{33}
\]

The second term in the bound of (30) is expanded as
\[
\langle \lambda_{i,t+1}, A_i(x^* - \theta_t) \rangle = \langle \lambda_{i,t+1}, A_i(x^* - x_t) \rangle + \langle \lambda_{i,t+1}, A_i(x_t - \theta_t) \rangle \\
= \langle \lambda_{i,t+1}, A_i(x^* - x_t) \rangle + \langle \lambda_{i,t+1} - \lambda_{i,t+1}, -r_i(x_t, y_{i,t}) \rangle + \langle \lambda_{i,t+1}, A_i(x_t - \theta_t) \rangle + \langle \lambda_{i,t+1} - \lambda_{i,t+1}, r_i(x_{i,t}, y_{i,t}) \rangle \\
= -h_{\lambda_{i,t+1}}(w, w^* + \lambda_{i,t+1}, A_i(x_t - \theta_t)) + \langle \lambda_{i,t+1} - \lambda_{i,t+1}, r_i(x_{i,t}, y_{i,t}) \rangle. \tag{34}
\]

Bounding the second term of (30), we have
\[
\langle \lambda_{i,t+1}, A_i(x_t - \theta_t) \rangle \leq \sigma_1(A_i) \| \lambda_{i,t+1} \| \| x_t - \theta_t \|. \tag{35}
\]

Moreover, applying (29) and (32) to (35), it follows that
\[
\sum_{t=1}^{T} \langle \lambda_{i,t+1}, A_i(x_t - \theta_t) \rangle \leq 2k \sqrt{T} \zeta \tag{36}
\]
Expanding the final term of (30) by applying line 7 of the algorithm and an inner product equality, we obtain\(^4\)
\[
\langle \lambda_{i,t+1}^*, r_i(x_{i,t}, y_{i,t}) \rangle = \frac{1}{\rho} \langle \lambda_{i,t+1}^*, \lambda_{i,t+1} - \lambda_{i,t} \rangle \\
= \frac{1}{2\rho} \left( \| \lambda_{i,t+1}^* - \lambda_{i,t} \|^2 + \| \lambda_{i,t} - \lambda_{i,t+1} \|^2 - 2 \| \lambda_{i,t} - \lambda_{i,t+1} \|^2 \right) \\
= \frac{1}{2\rho} \left( \| \lambda_{i,t}^* - \lambda_{i,t} \|^2 - \| \lambda_{i,t}^* - \lambda_{i,t+1} \|^2 \right) - \frac{\rho}{2} \| r_i(x_{i,t}, y_{i,t}) \|^2.
\]

Resolving the telescoping sum
\[
\sum_{t=1}^{T} \| \lambda_{i,t}^* - \lambda_{i,t+1} \|^2 - \| \lambda_{i,t}^* - \lambda_{i,t+1} \|^2,
\]
using the fact \( \lambda_{i,1} = 0, \) it now follows that
\[
\sum_{t=1}^{T} \langle \lambda_{i,t+1}^*, r_i(x_{i,t}, y_{i,t}) \rangle \leq \\
\frac{1}{2\rho} \left( \| \lambda_{i,t}^* \|^2 + \| \lambda_{i,t}^* - \lambda_{i,t+1} \|^2 \right) - \frac{\rho}{2} \sum_{t=1}^{T} \| r_i(x_{i,t}, y_{i,t}) \|^2 \\
\leq \frac{1}{2\rho} (2 \| \lambda_{i,t+1}^* \| \| \lambda_{i,t+1} \|) - \frac{\rho}{2} \sum_{t=1}^{T} \| r_i(x_{i,t}, y_{i,t}) \|^2.
\]

Applying (32) and the assumption \( \| \lambda_{i,t}^* \| \leq D_{\lambda} \)
\[
\sum_{t=1}^{T} \langle \lambda_{i,t+1}^*, r_i(x_{i,t}, y_{i,t}) \rangle \leq \\
\frac{D_{\lambda} \| \lambda_{i,t+1} \|}{\rho \sigma_1(A_i)} - \frac{\rho}{2} \sum_{t=1}^{T} \| r_i(x_{i,t}, y_{i,t}) \|^2. \tag{37}
\]

Substituting (36) and (37) into (34),
\[
\frac{1}{n} \sum_{i=1}^{n} \sum_{t=1}^{T} \langle \lambda_{i,t+1}, A_i(x^* - \theta_t) \rangle \\
\geq J_1 - \frac{1}{n} \sum_{i,t} \left[ h_{\lambda_{i,t+1}}(w, w^*) + \frac{\rho}{2} \| r_i(x_{i,t}, y_{i,t}) \|^2 \right] + 2k \sqrt{T} \times \tag{38}
\]
\[
(Q \zeta_\lambda + \zeta_{\text{max}}) + 2L_f \zeta + \zeta^2 + \zeta. \]

\textbf{Corollary 2. Given the sequence } \langle \psi(t) \rangle \textbf{ generated by Algorithm 1 with } \psi(x^*) \leq \Psi^2 \textbf{ and } \alpha(t) = k / \sqrt{T}, \textbf{ we have }
\[
R_T \leq J_1 + J_2 k \sqrt{T}, \tag{41}
\]

\textbf{The following corollary provides a more conservative bound on regret.}

\textbf{Corollary 2. Given the sequence } \langle \psi(t) \rangle \textbf{ generated by Algorithm 1 with } \psi(x^*) \leq \Psi^2 \textbf{ and } \alpha(t) = k / \sqrt{T}, \textbf{ we have }
\[
R_T \leq J_1 + J_2 k \sqrt{T}, \tag{41}
\]
where
\[ J_1 = \frac{3D_\lambda \zeta_{\text{max}}}{2 \rho \mu} \sum_{i=1}^{n} \frac{1}{\sigma_1(A_i)}, \]
\[ J_2 = (\zeta_{\text{max}} + L_f)^2(5 + 2Q) + \frac{\Psi^2}{k^2}. \]
\[ Q = \frac{\sqrt{n}}{1 - \sigma_2(P)}, \]
and \( \zeta_i = \sqrt{m} L_\phi \sigma_{m_i}(B_i^T) \) that is associated with local linear constraint at node \( i \). Moreover, \( \zeta_{\text{max}} = \max_i \zeta_i. \)

Proof: From Theorem 1 we have
\[ R_T \leq \tilde{J}_1 + \tilde{J}_2 k \sqrt{T}, \] (42)
where
\[ \tilde{J}_1 = \frac{3D_\lambda}{2 \rho \mu} \sum_{i=1}^{n} \frac{\zeta_i}{\sigma_1(A_i)}, \]
\[ \tilde{J}_2 = \zeta(4Q L_f + 2Q \zeta_{\text{max}} + 10L_f + 3\zeta) + 2\zeta + 5L_f^2 + 2Q L_f^2 + \frac{\Psi^2}{k^2}, \]
\[ \zeta = \frac{1}{n} \sum_{i=1}^{n} \zeta_i, \quad \tilde{\zeta} = \frac{1}{n} \sum_{i=1}^{n} \zeta_i^2, \quad \zeta_{\text{max}} = \max_i \zeta_i, \]
\[ \zeta_i = \sqrt{m} L_\phi \sigma_{1}(A_i) \sigma_{m_i}(B_i^T), \]
and \( Q = \frac{\sqrt{n}}{1 - \sigma_2(P)}. \)

Note that \( \zeta_i \leq \zeta_{\text{max}} \) which implies \( \tilde{J}_1 \leq J_1 \). Moreover, since \( \zeta \leq \zeta_{\text{max}} \) and \( \zeta \leq (\zeta_{\text{max}})^2 \) we have
\[ \tilde{J}_2 \leq \zeta_{\text{max}}(4Q L_f + 2Q \zeta_{\text{max}} + 10L_f + 5\zeta_{\text{max}}) + 5L_f^2 + 2Q L_f^2 + \frac{\Psi^2}{k^2}. \]

Thus, \( \tilde{J}_2 \leq J_2 \), and subsequently the statement of corollary follows.

The theorem and its corollary validate the “good” performance of OD-ADMM by demonstrating a sub-linear regret. In addition, it highlights the importance of the underlying interaction topology through \( \sigma_2(P) \) and the local linear constraints through \( \sigma_1(A_i) \) and \( \sigma_{m_i}(B_i) \). Further, a well known measure of network connectivity is the second smallest eigenvalue of the graph Laplacian \( L(G) \) denoted by \( \Lambda_2(G) \).

Since the communication matrix \( P \) is formed as proposed in Proposition 3, \( 1 - \sigma_2(P) \) is proportional to \( \Lambda_2(G) \) implying that high network connectivity promotes good performance of the proposed algorithm.

V. EXAMPLE - FORMATION ACQUISITION WITH POINTS OF INTEREST AND BOUNDARY CONSTRAINTS

Consider a formation acquisition problem amongst \( n \) agents where the position of agent \( i \), denoted as \( y_i \), is restricted to the convex set \( Y = [-1, 1]^2 \). The centroid of the formation is \( x \in \mathbb{R}^2 \) which is similarly constrained to \( X = Y \). The formation shape is defined for each agent by its offset \( c_i \) from the centroid, namely \( x - y_i = c_i \). There is a known boundary \( S \) which agents are required to avoid by increasing the distance to the boundary \( \text{dist}(y_i, S) = \inf_{x \in S} ||x - y_i|| \). This is achieved with a penalty function \( \phi_i(y_i) = (\text{dist}(y_i, S) + 1)^{-1} \) associated with agent \( y_i \)’s proximity to \( S \). Assuming that \( (S \cap \mathcal{X}) \) is an empty set, \( \phi_i(y_i) \) is convex. At each time step \( t \), each agent \( i \) obtains a location of interest \( q_{i,t} \) and the centroid is ideally located close to these locations of interest promoted through the minimization of the function \( f_{i,t}(x) = \frac{1}{2}||x - q_{i,t}||^2_2 \). The example illustrated in Figure 3 takes the form of problem (8), namely
\[
\min_{x \in \mathcal{X}, y_1,\ldots,y_n \in \mathcal{Y}} \sum_{t=1}^{T} \sum_{i=1}^{n} (f_{i,t}(x) + \phi_i(y_i))
\text{ s.t. } A_i x + B_i y_i = c_i \text{ for all } i \in [n],
\]
where \( A_i = -B_i = I_2 \) for all \( i \in [n] \).

Consider \( S = \{(x,y) \in \mathbb{R}||x| = 1.5, |y| = 1.5\} \) and so \( \phi_i(y_i) = (2.5 - ||y_i||^2_\infty)^{-1} \). The relevant parameters of the ADMM algorithm are \( g_i(t) = \nabla f_{i,t}(x_i) = x_i - q_{i,t}, k = 2, \rho = 0.5, \) and \( \psi(x) = ||x||^2_2 \). The remaining terms of the regret bound are \( L_\phi = 4/9, L_f = \sqrt{2}, \sigma_1(A_i) = \sigma_{m_i}(B_i) = 1, D_\lambda = 2, \) and \( K = 1 \).

The algorithm was applied to \( n = 8 \) agents connected over a random graph (see Figure 6) with \( \sigma_2(P) = 0.78 \) with \( c_i \)’s selected to acquire a formation with \( n \) agents equidistant apart on the circumference of a circle of radius 0.4. Locations of interest switch at each time step between a uniform distribution over the area of a length 0.5 square centered at \((-0.75, 0)\) and a Gaussian distribution with mean \((0, -0.75)\) and standard deviation 0.01, with bounds outside of \( \mathcal{X} \) ignored. The distribution is displayed in Figure 4. The sub-linear numerical regret, where \( w \) is the optimal solution of problem (8), compared to the theoretical regret appears in Figure 5a. The convergence of the global variables \( x_{i,t} \) to agreement as well as the reduction of the residue over time are displayed in Figure 5b.

The performance of the algorithm was compared for different graph topologies, namely path, star, cycle, random, cube and complete graph. These graph topologies are displayed in Figure 6. The matrix \( P \) was formed as proposed in Proposition 3, with \( \epsilon = d_{\text{max}} + 1 \), as such \( \sigma_2(P) = 1 - \frac{1}{2} \Lambda_2(L(G)) \). Under the same locations of interest as described previously the performance of the regret per time \( R_T/T \) for each graph topology is compared in Figure 7. The performance is strongly correlated to \( \sigma_2(P) \), as predicted in Theorem 1.
Figure 4: Distribution of locations of interest $q_{i,t}$.

Figure 5: (a) The regret per time $R_{T}/T$ compared with its theoretical bound in Theorem 1. (b) The standard deviation of the global variable $x_i$ and the average residue for each agent over times smoothed by taking the maximum over a $t_w = 1000$ sliding window.

Figure 6: Topologies of the 6 different graph types.

Figure 7: The regret per time $R_{T}/T$ performance of 6 different graph types, specifically a path, star, cycle, random, cube and complete graph with $\sigma_2(P) = \{0.95, 0.88, 0.80, 0.78, 0.50, 0.00\}$, respectively. The trajectories are smoothed by taking the average over a $t_w = 1000$ sliding window.

VI. CONCLUSION

A new problem set up was introduced with a network of agents, where each agent optimizes the global objective function with access to its privately known local objective function and linear constraint. This formulation has many applications in networked systems such as robotics and computer networks. A decentralized algorithm allows us to make the decisions parallel in time based on local information and communication with neighboring agents.

The fully decentralized online algorithm, developed in this paper, achieves a sub-linear regret bound of $O(\sqrt{T})$ for the objective function and linear local constraints violation. In particular, this algorithm is competitive with respect to the best fixed decision performance in hindsight. Moreover, we highlight the role of the underlying network topology in achieving “good” regret, i.e., the regret bound improves with increased connectivity in the network.

The proposed algorithm was applied to a formation acquisition problem showing agreement with the theoretical results. Future work of particular interest includes exploring regret bound over a time varying network topology, and...
investigating favorable graph characteristics for the online ADMM framework.

VII. APPENDIX

The following results can be found in [9] and [10], respectively. Thus, they are presented here with absent or abridged proofs.

**Proposition 3.** If graph $G$ is strongly connected then the matrix $P = I - \frac{1}{c} \text{diag}(v)L(G)$ is doubly stochastic, where $v^T L(G) = 0$ with positive vector $v = [v_1, v_2, \ldots, v_n]^T$ and $c \in (\max_{i \in V} (v_d_v), \infty)$. If graph $G$ is balanced then the matrix $P = I - \frac{1}{c} L(G)$ is doubly stochastic, where $c \in (d_{\text{max}}, \infty)$.

**Lemma 4.** For any $u, v \in \mathbb{R}^m$, and under the conditions stated for proximal function $\psi$ and step size $\alpha$, we have

$$\|\prod_{\chi}(u, \alpha) - \prod_{\chi}(v, \alpha)\| \leq \alpha \|u - v\|_\ast.$$  

**Lemma 5.** For any positive and non-increasing sequence $\alpha(t)$ and $x^* \in \chi$

$$\sum_{t=1}^{T} (g_t, \theta_t - x^*) \leq \sum_{t=1}^{T} (A^T \lambda_{i,t+1}, x^* - \theta_t) + \frac{1}{\alpha t} \psi(x^*) + \sum_{t=2}^{T} \frac{\alpha_{t-1}}{2} \|g_t + A^T \lambda_{i,t+1}\|^2_\ast,$$

where the sequence of $\theta_t$ is generated by (23)-(24).

**Proof:** Based on Lemma 3 in [9], we have

$$\sum_{t=1}^{T} (g_t, A^T \lambda_{i,t+1}, \theta_t - x^*) \leq \sum_{t=1}^{T} \frac{\alpha_{t+1}}{2} \|g_t + A^T \lambda_{i,t+1}\|^2_\ast,$$

and the statement of the lemma follows.

**Lemma 6.** For any sequences of $z_{i,t}$ and $z_t$ generated by Algorithm 1, we have

$$\|z_t - z_{i,t}\|_\ast \leq \sqrt{n} (L_f + \epsilon_{\text{max}}) \frac{1}{1 - \sigma_2(P)} + 2L_f + \bar{\zeta} + \bar{\zeta}_i$$

for all $i \in [n]$ and $t \in [T]$, where $\bar{\zeta}_i = L_f \sigma_1(A)$, $\bar{\zeta} = \frac{1}{n} \sum_{i=1}^{n} \zeta_i$ and $\epsilon_{\text{max}} = \max_{i} \zeta_i$.

**Proof:** Based on line 8 of Algorithm 1 we have

$$z_{i,t} = \sum_{j=1}^{n} [P^s]_{ji} z_{j,t-s} + g_{i,t-1} + A^T \lambda_{i,t+1} + \sum_{k=t-1}^{t-2} \sum_{j=1}^{n} [P^{t-k-1}]_{ji} (g_{j,k} + A^T \lambda_{j,k+1}).$$

In addition, $z_t$ evolves as

$$z_t = z_{t-s} + \sum_{k=t-s}^{t-1} \sum_{j=1}^{n} \frac{1}{n} (g_{j,k} + A^T \lambda_{j,k+1}).$$  

Assuming $t - s = 1$, and $z_{i,1} = 0$ for all $i \in [n]$ and based on (43) we have

$$z_t - z_{i,t} = \sum_{k=1}^{t-2} \sum_{j=1}^{n} \frac{1}{n} - [P^t-k-1]_{ji} (g_{j,k} + A^T \lambda_{j,k+1}) + \frac{1}{n} \sum_{j=1}^{n} A^T \lambda_{j,t} - A^T \lambda_{i,t} + g_{i,t-1} - g_{i,t-1}.$$

Thus, the dual norm of $z_t - z_{i,t}$ can be bounded as

$$\|z_t - z_{i,t}\|_\ast \leq \sum_{k=1}^{t-2} \sum_{j=1}^{n} \|g_{j,k} + A^T \lambda_{j,k+1}\|_\ast \frac{1}{n} - [P^t-k-1]_{ji} + \frac{1}{n} \sum_{j=1}^{n} \|A^T \lambda_{j,t}\|_\ast + \|A^T \lambda_{i,t}\|_\ast + \|g_{i,t-1} - g_{i,t-1}\|_\ast$$

$$\leq \sum_{j=1}^{t-2} \max_{k} \|g_{j,k} + A^T \lambda_{j,k+1}\|_\ast \|P^t-k-1\|_\ast + \frac{1}{n} \sum_{j=1}^{n} \sigma_1(A) \|\lambda_{j,t}\|_\ast + \|g_{i,t-1} - g_{i,t-1}\|_\ast.$$

Since $\|g_{i,t}\|_\ast \leq L_f$ and $\|A^T \lambda_{i,t}\|_\ast \leq \bar{\zeta}_i \leq \epsilon_{\text{max}}$, the dual norm of $z_t - z_{i,t}$ is further bounded as

$$\|z_t - z_{i,t}\|_\ast \leq \sqrt{n} (L_f + \epsilon_{\text{max}}) \frac{1}{1 - \sigma_2(P)} + 2L_f + \frac{1}{n} \sum_{j=1}^{n} \zeta_j.$$

In addition, the second largest singular value of $P$ is $\sigma_2(P) \leq 1$, as $P$ is a doubly stochastic matrix [22]. Thus, the inequality (46) is bounded as

$$\|z_t - z_{i,t}\|_\ast \leq \sqrt{n} (L_f + \epsilon_{\text{max}}) \frac{1}{1 - \sigma_2(P)} + 2L_f + \bar{\zeta} + \bar{\zeta}_i.$$

REFERENCES


\footnote{Note that $\|P^tx - \frac{1}{n} x\|_\ast \leq \sigma_2(P) \sqrt{n}$, where the vector $x$ belongs to $x \in \mathbb{R}^n | x \in \mathbb{R}^n, \sum_{i=1}^{n} x_i = 1$. This is a property of stochastic matrix $P$ introduced by Duchi et al. [9].}


