

DISTRIBUTED OPTIMIZATION FOR EVOLVING NETWORKS OF GROWING CONNECTIVITY

Sijia Liu Pin-Yu Chen Alfred O. Hero III

Department of EECS, University of Michigan, Ann Arbor, MI 48109, USA
{lsjxjtu, pinyu, hero}@umich.edu

Abstract—We focus on the problem of distributed optimization for multi-agent networks via distributed dual averaging (DDA) over an evolving network of growing connectivity. It is known that the convergence rate of DDA is influenced by the algebraic connectivity of the underlying network, where better connectivity leads to faster convergence. However, the effect of the growth of network connectivity on the convergence rate has not been fully understood. This paper provides a tractable approach to analyze the improvement in the convergence rate of DDA induced by the growth of network connectivity. This analysis is applicable, for example, to successive refinement strategies in massive multi-core optimizers where an increasing number of local data passage edges are successively added between cores in order to accelerate total run time. Compared to the existing convergence results, our analysis gives tighter bounds on the convergence of DDA over networks of growing connectivity. Numerical experiments show that our analysis leads to orders of improvement for evaluating convergence rate, which is not captured by existing analysis.

Keywords—Distributed optimization, dual averaging, graph Laplacian, growing connectivity, multi-agent network.

I. INTRODUCTION

In recent years, distributed optimization has received extensive attention. Network-structured optimization problems have found a wide range of applications in parallel computing [1], [2], sensor networks [3]–[5], and power grids [6]. Many such problems can be formulated as a constrained minimization problem, in which the network loss function is given as a sum of local objective functions accessed by agents, e.g., cores in massive parallel computing. The goal of distributed optimization is to find the solution of the network-structured optimization problem using only local computation and communication at each agent.

There has been a significant amount of research on developing distributed computation and optimization algorithms in multi-agent systems. Early research efforts focus on the design of consensus or diffusion algorithms for distributed averaging, estimation and filtering [7]–[14]. Triggered by diffusion adaptation and consensus learning, distributed subgradient algorithms were proposed under various scenarios, e.g., networks with communication delays, dynamic topologies, link failures, and push-sum averaging protocols [15]–[17]. In this paper, we focus on a distributed dual averaging (DDA) subgradient algorithm [18], where the key ingredient of DDA is to maintain a weighted average of subgradients throughout the network.

Recently, DDA-based subgradient algorithms have attracted great attention in signal processing over networks, machine learning, parallel computation and online convex optimization [18]–[23]. In [18], the convergence analysis of DDA was firstly developed for distributed optimization over static networks. It was shown that there exists a tight connection between the convergence rate of DDA and the algebraic connectivity (namely, the second-smallest Laplacian eigenvalue [24]) of the underlying network topology. Although the convergence of DDA with stochastic and time-varying communications has been discussed in [18], such a convergence analysis assumes confined

variations in network connectivity for dynamic networks such that the convergence results developed for static networks can be applied. As a result, in this paper we refer to the convergence analysis in [18] as the *static network* approach.

The convergence of DDA over networks with time-varying topologies was investigated in recent work [22], [23]. Given an arbitrary time-varying sequence of network topologies, it is shown in [22] that the convergence rate of DDA is loosely bounded by the convergence of the least connected network in the sequence. To further understand the effect of dynamic network connectivity on the convergence of DDA, we focus on the convergence analysis of DDA over networks of growing connectivity. The case of growing connectivity is inspired by real-world scenarios. One compelling example is in adaptive mesh parallel scientific computing where the network corresponds to grid points in the mesh and optimization is performed by solving partial differential equations over a domain that is successively and adaptively refined over the network as time progresses [25]. As another example, the accumulated connectivity of online social networks increases over time when users establish new connections (e.g., time-evolving friendship in Facebook or LinkedIn). Another example arises in the design of resilient hardened physical networks [26]–[29], where adding edges or rewiring existing edges increases network connectivity and robustness to node or edge failures. Based on that, distributed optimization can be performed over a sparse network of low computation and communication cost in the beginning, and then over a sequence of well-designed networks for improved convergence.

In this paper, we provide a novel convergence analysis of DDA over networks with growing connectivity. There exist two major contributions. First, we devise new methods for analyzing the effect of network topologies with growing connectivity on the convergence rate of DDA. Second, we provide a tractable approach to quantify the improvement in convergence rate induced by the growth of network connectivity. Extensive numerical results show an excellent agreement between the empirical convergence behavior and our theoretical predictions. Compared to the existing convergence results, our analysis leads to improved characterization of convergence rates.

II. PRELIMINARIES: GRAPH, DISTRIBUTED OPTIMIZATION AND DISTRIBUTED DUAL AVERAGING

In this section, we provide background on graphical models for multi-agent networks, the distributed optimization problem, and the distributed dual averaging algorithm.

A. Graphical model for multi-agent networks

A graph yields a succinct representation of interactions among agents or sensors over a network. Let $\mathcal{G}_t = (\mathcal{V}, \mathcal{E}_t)$ denote a *time-varying* undirected unweighted graph, where \mathcal{V} is a node set with cardinality $|\mathcal{V}| = n$, and $\mathcal{E}_t \subseteq [n] \times [n]$ is an edge set at time t . For simplicity, we denote by $[n]$ the integer set $\{1, 2, \dots, n\}$. An edge $(i, j) \in \mathcal{E}_t$ indicates that there exists a communication link between

This work was partially supported by grants from the US Army Research Office, grant numbers W911NF-15-1-0479 and W911NF-15-1-0241.

agent i and agent j at time t . And a neighborhood of agent i is given by $\mathcal{N}_i(t) = \{j \mid (i, j) \in \mathcal{E}_t\}$.

A graph can be represented equivalently by an adjacency matrix or a graph Laplacian matrix. Let \mathbf{A}_t be the adjacency matrix of \mathcal{G}_t , where $[\mathbf{A}_t]_{ij} = 1$ for $(i, j) \in \mathcal{E}_t$ and $[\mathbf{A}_t]_{ij} = 0$, otherwise. Here $[\mathbf{X}]_{ij}$ (or \mathbf{X}_{ij}) denotes the (i, j) -th entry of a matrix \mathbf{X} . The graph Laplacian matrix is defined as $\mathbf{L}_t = \mathbf{D}_t - \mathbf{A}_t$, where \mathbf{D}_t is a degree matrix, whose i -th diagonal entry is given by $\sum_j [\mathbf{A}_t]_{ij}$. The Laplacian matrix is always positive semidefinite, and has a zero eigenvalue $\lambda_n = 0$ (eigenvalues are sorted in decreasing order of magnitude) with eigenvector $(1/\sqrt{n})\mathbf{1}$, where $\mathbf{1}$ is the column vector of ones. The second-smallest Laplacian eigenvalue λ_{n-1} is known as the algebraic connectivity [24], which is positive if and only if the graph is connected, namely, there exists a communication path between every pair of distinct nodes.

In this paper, we assume that each \mathcal{G}_t is connected, and the resulting algebraic connectivity $\lambda_{n-1}(\mathbf{L}_t)$ is monotonically increasing over time, that is,

$$0 < \lambda_{n-1}(\mathbf{L}_0) \leq \lambda_{n-1}(\mathbf{L}_1) \leq \dots \leq \lambda_{n-1}(\mathbf{L}_T), \quad (1)$$

where T is the length of time horizon.

B. Distributed optimization

We consider a convex optimization problem based on local cost functions, each of which is associated with a node/agent. The objective is to minimize the average cost over the network,

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) := \frac{1}{n} \sum_{i=1}^n f_i(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{x} \in \mathcal{X}, \end{aligned} \quad (2)$$

where $\mathbf{x} \in \mathbb{R}^d$ is the optimization variable, f_i is convex and L -Lipschitz continuous¹, and \mathcal{X} is a closed convex set containing the origin. A concrete example of (2) is a distributed estimation problem, where f_i is a square loss and \mathbf{x} is an unknown parameter to be estimated. A graph \mathcal{G}_t imposes communication constraints of distributed optimization. That is, each node i only accesses to the local cost function f_i and can communicate directly only with nodes in its neighborhood $\mathcal{N}_i(t)$ at time t .

C. Distributed dual averaging (DDA)

Throughout this paper we employ the distributed dual averaging algorithm [18] to solve the optimization problem (2) in a decentralized manner. To be specific, each node $i \in \mathcal{V}$ performs the updates

$$\mathbf{z}_i(t+1) = \sum_{j \in \mathcal{N}_i(t)} [\mathbf{P}_t]_{ji} \mathbf{z}_j(t) + \mathbf{g}_i(t) \quad (3)$$

$$\mathbf{x}_i(t+1) = \arg \min_{\mathbf{x} \in \mathcal{X}} \left\{ \mathbf{z}_i(t+1)^T \mathbf{x} + \frac{1}{\alpha_t} \psi(\mathbf{x}) \right\}, \quad (4)$$

where $\mathbf{z}_i(t) \in \mathbb{R}^d$ is an auxiliary variable for node i at time t , $\mathbf{P}_t \in \mathbb{R}^{n \times n}$ is a matrix of non-negative weights that preserves the zero structure of the graph Laplacian \mathbf{L}_t , $\mathbf{g}_i(t)$ is a subgradient of $f_i(\mathbf{x})$ at $\mathbf{x} = \mathbf{x}(t)$, $\psi(\mathbf{x})$ is a regularizer for stabilizing the update, and $\{\alpha_t\}_{t=0}^{\infty}$ is a non-increasing sequence of positive step-sizes. In (4), $\psi(\mathbf{x})$ is also known as a proximal function, which is assumed to be 1-strongly convex with respect to a generic norm $\|\cdot\|$, and $\psi(\mathbf{x}) > 0$ and $\psi(\mathbf{0}) = 0$. In particular, when $\|\cdot\|$ is the ℓ_2 norm, we obtain the canonical proximal function $\psi(\mathbf{x}) = (1/2)\|\mathbf{x}\|_2^2$.

¹The L -Lipschitz continuity of f with respect to a generic norm $\|\cdot\|$ is defined by $|f(\mathbf{x}) - f(\mathbf{y})| \leq L\|\mathbf{x} - \mathbf{y}\|$, for $\mathbf{x}, \mathbf{y} \in \mathcal{X}$.

The weight matrix \mathbf{P}_t in (3) is assumed to be doubly stochastic, namely, $\mathbf{1}^T \mathbf{P}_t = \mathbf{1}^T$ and $\mathbf{P}_t \mathbf{1} = \mathbf{1}$. A common choice of \mathbf{P}_t that associates with the graph structure is given by

$$\mathbf{P}_t = \mathbf{I} - \frac{1}{2(1 + \delta_{\max,t})} \mathbf{L}_t, \quad (5)$$

where $\delta_{\max,t}$ is the maximum degree of \mathcal{G}_t . Although many choices of \mathbf{P}_t are possible, it is often the case that \mathbf{P}_t is constructed by \mathbf{L}_t [14], [18], [22], [23]. The considered \mathbf{P}_t in (5) is corresponding to a lazy random walk and is positive semidefinite [30]. For networks of growing connectivity, from (1) and (5) we obtain

$$\sigma_2(\mathbf{P}_0) \geq \sigma_2(\mathbf{P}_1) \geq \dots \geq \sigma_2(\mathbf{P}_T), \quad (6)$$

where $\sigma_2(\mathbf{P}_t)$ is the second-largest singular value of \mathbf{P}_t .

III. MAIN RESULTS: CONVERGENCE ANALYSIS

In this section, we establish a theoretical connection between the growing connectivity and the convergence rate of DDA. It is known from [18] that for each agent $i \in [n]$, the convergence of the running local average $\bar{\mathbf{x}}_i(T) = (1/T) \sum_{t=1}^T \mathbf{x}_i(t)$ to the solution of problem (2), denoted by \mathbf{x}^* , is governed by two error terms: a) optimization error common to subgradient algorithms, and b) network penalty due to the cost of node communications. We summarize the basic convergence result in Theorem 1.

Theorem 1 [18, Theorem 1]: Given the updates (3) and (4), the difference $f(\bar{\mathbf{x}}_i(T)) - f(\mathbf{x}^*)$ for $i \in [n]$ is upper bounded as $f(\bar{\mathbf{x}}_i(T)) - f(\mathbf{x}^*) \leq \text{OPT} + \text{NET}$. Here

$$\text{OPT} = \frac{1}{T\alpha_T} \psi(\mathbf{x}^*) + \frac{L^2}{2T} \sum_{t=1}^T \alpha_{t-1}, \quad (7)$$

$$\text{NET} = \sum_{t=1}^T \frac{L\alpha_t}{T} \left(\frac{2}{n} \sum_{j=1}^n \|\bar{\mathbf{z}}(t) - \mathbf{z}_j(t)\|_* + \|\bar{\mathbf{z}}(t) - \mathbf{z}_i(t)\|_* \right), \quad (8)$$

where $\bar{\mathbf{z}}(t) = (1/n) \sum_{i=1}^n \mathbf{z}_i(t)$, and $\|\cdot\|_*$ is the dual norm² to $\|\cdot\|$.

Note that the optimization error (7) can be made arbitrarily small for a sufficiently large T and an appropriate α_t , e.g., $\alpha_t \propto 1/\sqrt{t}$. The network penalty (8) measures the deviation of each node's local estimate from the average consensus value. In what follows, we will bound (8) under the condition (6), induced by the increasing connectivity.

Let $\Phi(t, s)$ denote the product of time-varying stochastic matrices, namely, $\Phi(t, s) = \mathbf{P}_t \mathbf{P}_{t-1} \times \dots \times \mathbf{P}_s$, where $s \leq t$. To bound (8), we begin by relating $\sigma_2(\Phi(t, s))$ to $\{\sigma_2(\mathbf{P}_t)\}$. This is formally stated as a lemma.

Lemma 1: Given $\Phi(t, s) = \mathbf{P}_t \mathbf{P}_{t-1} \times \dots \times \mathbf{P}_s$, we obtain

$$\sigma_2(\Phi(t, s)) \leq \prod_{i=s}^t \sigma_2(\mathbf{P}_i), \quad (9)$$

where $\sigma_2(\mathbf{M})$ is the second largest singular value of a matrix \mathbf{M} .

Proof: The proof is reported in [31, Appendix A]. ■

Based on Theorem 1 and Lemma 1, Proposition 1 below shows that the upper bound on the network penalty (8) is controlled by the spectral gap $(1 - \sigma_2(\mathbf{P}_0))$ and its temporal variation given by (6).

² $\|\mathbf{v}\|_* := \sup_{\|\mathbf{u}\|=1} \mathbf{v}^T \mathbf{u}$.

Proposition 1: Under the growing connectivity condition (6) and $\mathbf{z}_i(0) = \mathbf{0}$ for $i \in [n]$, the network penalty (8) is bounded as

$$\text{NET} \leq \sum_{t=1}^T \frac{L^2 \alpha_t}{T} \left(6 \left[\frac{\log T \sqrt{n}}{1 - \sigma_2(\mathbf{P}_0)} - \frac{\log \beta^{*-1}}{1 - \sigma_2(\mathbf{P}_0)} \right] + 9 \right), \quad (10)$$

where $\lceil x \rceil$ gives the smallest integer that is greater than x , and β^* is the solution of the optimization problem

$$\underset{\beta}{\text{minimize}} \quad \delta := \left[\frac{\log T \sqrt{n}}{\log \sigma_2(\mathbf{P}_0)^{-1}} - \frac{\log \beta^{-1}}{\log \sigma_2(\mathbf{P}_0)^{-1}} \right] \quad (11a)$$

$$\text{subject to} \quad \prod_{i=0}^{\delta-1} \sigma_2(\mathbf{P}_i) \leq \beta \sigma_2(\mathbf{P}_0)^\delta \quad (11b)$$

$$\frac{1}{T \sqrt{n}} < \beta \leq 1. \quad (11c)$$

Proof: The proof is reported in [31, Appendix B]. ■

Before delving into interpreting Proposition 1, we elaborate on the optimization problem (11). Here the variable β describes the temporal variation of $\sigma_2(\mathbf{P}_t)$ against $\sigma_2(\mathbf{P}_0)$, and δ denotes the objective value. The objective function (11a) implies that $\delta \geq \log(\beta T \sqrt{n}) / \log \sigma_2(\mathbf{P}_0)^{-1}$, namely, $\beta \sigma_2(\mathbf{P}_0)^\delta \leq 1 / (T \sqrt{n})$. One can understand that δ quantifies the temporal mixing time incurred by networks of growing connectivity so that $\prod_{i=0}^{\delta-1} \sigma_2(\mathbf{P}_i) \leq 1 / (T \sqrt{n})$. The constraint (11b) implies that the more the connectivity of a network grows, the smaller β becomes. In (11c), the lower bound on β stems from $\delta \geq 1$, and the upper bound is due to the growth of network connectivity. We finally remark that $\beta = 1$ is a feasible point to problem (11), and the optimal β is achieved by searching for the interval $(1/T \sqrt{n}, 1]$ until δ given by (11a) is minimized and the inequality (11b) is satisfied.

Proposition 1 reveals a tight connection between the convergence rate of DDA and the spectral properties of the time-varying network through β^* and $1 - \sigma_2(\mathbf{P}_0)$. For example, if the network is static, namely, $\sigma_2(\mathbf{P}_t) = \sigma_2(\mathbf{P}_0)$ for $t \in [T]$, we obtain $\beta^* = 1$ from (11), and the right hand side of (10) reduces to the error bound proposed by [18]. Based on (5), the spectral gap $1 - \sigma_2(\mathbf{P}_0)$ can be easily associated with the algebraic connectivity $\lambda_{n-1}(\mathbf{L}_0)$. Combining Theorem 1 and Proposition 1, we present the convergence rate of DDA over networks of growing connectivity in Theorem 2.

Theorem 2: Under the hypotheses of Theorem 1, $\psi(\mathbf{x}^*) \leq R^2$, and $\alpha_t \propto R \sqrt{1 - \sigma_2(\mathbf{P}_0)} / (L \sqrt{t})$, we obtain for $i \in [n]$

$$f(\hat{\mathbf{x}}_i(T)) - f(\mathbf{x}^*) = \mathcal{O} \left(\frac{RL}{\sqrt{T}} \left[\frac{\log T \sqrt{n}}{\sqrt{\lambda_{n-1}(\mathbf{L}_0)}} - \frac{\log \beta^{*-1}}{\sqrt{\lambda_{n-1}(\mathbf{L}_0)}} \right] \right), \quad (12)$$

where $f = \mathcal{O}(g)$ means that f is bounded above by g up to some constant factor, and β^* is the solution of problem (11). ■

It is clear from (12) that the term $\log \beta^{*-1} / \sqrt{\lambda_{n-1}(\mathbf{L}_0)}$ is introduced due to the growth of connectivity, where β decreases as the connectivity grows at a faster rate. When $\beta^* = 1$, the function error (12) reduces to that of [18] when the network is static. That is, $f(\hat{\mathbf{x}}_i(T)) - f(\mathbf{x}^*) = \mathcal{O} \left(\frac{RL}{\sqrt{T}} \frac{\log T \sqrt{n}}{\sqrt{\lambda_{n-1}(\mathbf{L}_0)}} \right)$.

IV. TOPOLOGY SWITCHING VERSUS CONVERGENCE TIME

To study the effect of the growth of network connectivity on the convergence time of DDA, in this section we introduce a topology switching rate to characterize the dynamics of the graph Laplacian

matrices of $\{\mathcal{G}_t\}$. The time-varying graph Laplacian matrix is specified as follows.

$$\mathbf{L}_t = \begin{cases} \mathbf{L}_0 & t \in [0, \Delta] \\ \mathbf{L}_1 & t \in [\Delta + 1, 2\Delta] \\ \vdots & \vdots \\ \mathbf{L}_q & t \in [q\Delta + 1, T], \end{cases} \quad (13)$$

where $\lambda_{n-1}(\mathbf{L}_{i+1}) > \lambda_{n-1}(\mathbf{L}_i)$ for $i \in [q]$, and Δ is the length of time interval during which the graph Laplacian matrix remains unaltered. In (13), the quantity $1/\Delta$ gives the topology switching rate, where the smaller Δ is, the faster the connectivity increases. In an extreme case of $\Delta = T$, the network becomes static with $\mathbf{L}_t = \mathbf{L}_0$ for $t \in [T]$.

Using (11a), the convergence rate given by Theorem 2 becomes

$$f(\hat{\mathbf{x}}_i(T)) - f(\mathbf{x}^*) = \mathcal{O} \left(\frac{RL}{\sqrt{T}} \delta^* \right), \quad i \in [n]. \quad (14)$$

where $\delta^* := \left[\frac{\log T \sqrt{n}}{\sqrt{\lambda_{n-1}(\mathbf{L}_0)}} - \frac{\log \beta^{*-1}}{\sqrt{\lambda_{n-1}(\mathbf{L}_0)}} \right]$ is directly proportional to Δ , since a small Δ (fast growing of network connectivity) leads to a fast mixing Markov chain that yields small β^* and δ^* . For ease of analysis, we assume that $\delta^* = \Delta^\tau$, where the convergence of DDA at $\Delta = T$ implies $\tau < 0.5$ from (14). We then obtain from (14) that at most $\mathcal{O}(\Delta^{2\tau}/\epsilon^2)$ iterations are required to achieve an ϵ -accurate solution. Compared to the convergence time under a static topology derived in [18], Proposition 2 shows a relative gain for the improvement in convergence time under the switching topology model (13). Moreover, extensive numerical results in Sec. V show that the empirical convergence time is well aligned with the theoretical prediction, which is not captured by existing convergence analysis.

Proposition 2: Let T_s and T_d denote the number of iterations required to achieve an ϵ -accurate solution under a static topology \mathbf{L}_0 and a time-varying topology \mathbf{L}_t in (13), respectively. If $\delta^* = \Delta^\tau$, the relative gain for the improvement in convergence time is given by $(T_s - T_d)/T_s = 1 - \Delta^{2\tau}/T_s^{2\tau}$.

Proof: Based on the analysis in [18], it is known that $T_s = C_s/\epsilon^2$, where C_s is a constant independent of ϵ . Based on (14) and $\delta^* = \Delta^\tau$, we have $T_d = C_d \Delta^{2\tau}/\epsilon^2$, where C_d is a constant independent of Δ and ϵ . Since $T_s = C_s/\epsilon^2 = C_d T_s^{2\tau}/\epsilon^2$ when $\Delta = T_s$, we obtain $C_s/C_d = T_s^{2\tau}$. And thus the relative gain is $(T_s - T_d)/T_s = 1 - \Delta^{2\tau} C_d/C_s = 1 - \Delta^{2\tau}/T_s^{2\tau}$. ■

Proposition 2 implies that $0 \leq (T_s - T_d)/T_s \leq (1 - 1/T_s^{2\tau})$, where the left hand side is achieved when $\Delta = T_s$, and the right hand side is achieved when $\Delta = 1$. As a result, our analysis explicitly characterizes the relation between growing network connectivity and convergence of DDA, whereas the existing convergence analysis (i.e., T_s) in [18] is insensitive to networks of growing connectivity. Moreover, our analysis shows that the improvement in convergence time can be significant if the network connectivity increases rapidly (i.e., the case of small Δ).

V. NUMERICAL RESULTS

In this section, numerical experiments are conducted to validate the above theory on convergence behavior of DDA over networks with growing connectivity. We will show that the empirical convergence behavior matches our theoretical predictions.

To specify a distribution optimization problem of the form in (2), we consider an ℓ_1 regression loss function $f_i(\mathbf{x}) = |y_i - \mathbf{b}_i^T \mathbf{x}|$ for $i \in [n]$, and $\mathcal{X} = \{\mathbf{x} \in \mathbb{R}^d \mid \|\mathbf{x}\|_2 \leq R\}$, where $\{y_i\}$ and $\{\mathbf{b}_i\}$ are data

points drawn from normal distribution, $n = 100$, $d = 5$, and $R = 5$. We note that f_i is L -Lipschitz continuous, where $L = \max_i \|\mathbf{b}_i\|_2$. For the underlying network model, we consider both k -regular ring network and random geometric graph [32]. The examples are given in Fig. 1. For a k -regular ring network, the connectivity grows by increasing k with respect to the switching topology model in (13). For a random geometric graph, since any two nodes separated by a distance less than some radius $r > 0$ are connected, we can increase r to obtain networks of growing connectivity with respect to (13). In the distributed dual averaging algorithm, we set \mathbf{P}_t as in (5), and choose stepsizes $\alpha_t \propto 1/\sqrt{t}$.

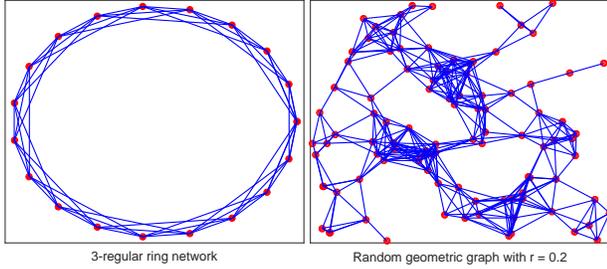


Fig. 1: Examples of 3-regular ring network and random geometric graph with $r = 0.2$ in a unit square region.

In Fig. 2, we present the function error $\max_i [f(\hat{\mathbf{x}}_i(t)) - f(\mathbf{x}^*)]$ versus the iteration index $t \in [T]$ for both k -regular ring network and random geometric graph with a varying Δ , where $T = 4000$. We recall from (13) that the parameter Δ governs the growing speed of network connectivity. For comparison, we also plot the convergence trajectory under the static network topology assumption in [18], which is a special case of our analysis when $\Delta = T$. As we can see, when Δ is small (namely, the connectivity increases fast), the convergence rate is significantly improved. This result is consistent with the theoretical implications in Theorem 2 and Proposition 2. Even with a relatively large Δ (say $\Delta = 500$), the convergence performance improves as compared to the case of static network topology.

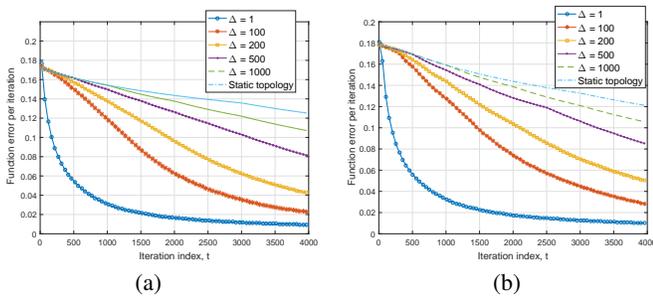


Fig. 2: Function error versus iterations for different values of Δ : a) k -regular ring network, and b) random geometric graph. These trends are consistent with the predictions of Theorem 2 and Proposition 2.

In Fig. 3, we present the function error $\max_i [f(\hat{\mathbf{x}}_i(T)) - f(\mathbf{x}^*)]$ ($T = 4000$) versus the value of Δ for both k -regular ring network and random geometric graph. We also plot the predicted function error given by (12) (scaled up to constant factor). As we can see, the function error decreases as Δ decreases due to the benefits of successively increased network connectivity. Further, we observe an excellent agreement between empirical function errors and theoretical predictions (Theorem 2) in all cases.

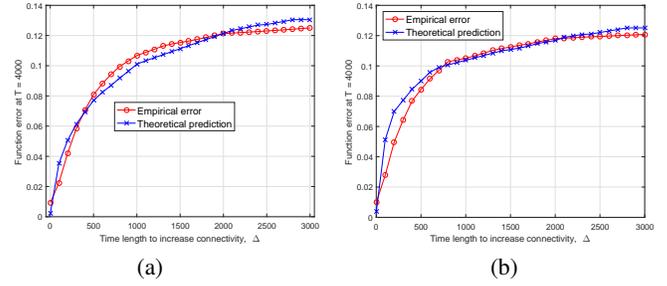


Fig. 3: Empirical and predicted function error at time T versus Δ : a) k -regular ring network, and b) random geometric graph. These trends are consistent with the predictions of Theorem 2.

Our final set of experiments investigates the convergence time of distributed dual averaging for k -regular ring network with increasing connectivity. In Fig. 4-(a), we present the convergence time, in terms of the number of iterations required to achieve $\max_i [f(\hat{\mathbf{x}}_i(t)) - f(\mathbf{x}^*)] \leq 0.1$, as a function of Δ . Compared to the convergence time under the static network assumption in [18], we observe a significant improvement on convergence time induced by the growth of network connectivity. In Fig. 4-(b), we compare the empirical improvement on convergence time to the theoretical prediction in Proposition 2 with $\tau = 0.175$. We observe that the empirical behavior is consistent with our theoretical predictions.

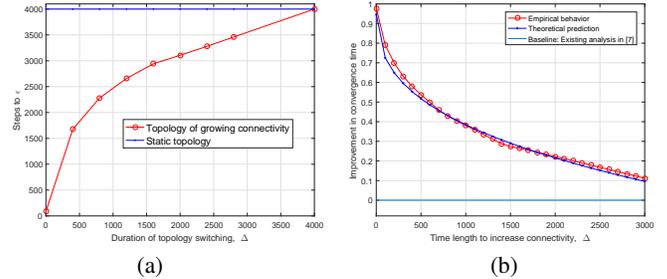


Fig. 4: Convergence time for k -regular ring network: a) number of iterations, and b) improvement compared to static topology. These trends are consistent with the predictions of Proposition 2.

VI. CONCLUSIONS

In this paper we have studied the distributed optimization problem in a dynamic multi-agent network whose algebraic connectivity grows over time. In this scenario, we have provided a novel convergence analysis of distributed dual averaging that is commonly used to solve the network-structured optimization problem. We have established a tight connection between the improvement in convergence rate of the algorithm and the growth speed of network connectivity. An excellent agreement between the empirical convergence behavior and our theoretical predictions has been shown via numerical results, which are not captured by the previous convergence analysis. For future work, we would like to relax the assumption of networks with growing connectivity, and to study the convergence rate of distributed optimization methods under both time-varying networks and connectivity cost constrains.

REFERENCES

- [1] B. Hendrickson and T. G. Kolda, "Graph partitioning models for parallel computing," *Parallel computing*, vol. 26, no. 12, pp. 1519–1534, 2000.
- [2] A. Pothen, "Graph partitioning algorithms with applications to scientific computing," in *Parallel Numerical Algorithms*, pp. 323–368. Springer, 1997.
- [3] D. Blatt, A. O. Hero, and H. Gauchman, "A convergent incremental gradient method with a constant step size," *SIAM Journal on Optimization*, vol. 18, no. 1, pp. 29–51, 2007.
- [4] R. Olfati-Saber, J. A. Fax, and R. M. Murray, "Consensus and cooperation in networked multi-agent systems," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 215–233, Jan. 2007.
- [5] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.
- [6] F. Dörfler, M. Chertkov, and F. Bullo, "Synchronization in complex oscillator networks and smart grids," *Proceedings of the National Academy of Sciences*, vol. 110, no. 6, pp. 2005–2010, 2013.
- [7] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Systems and Control Letters*, vol. 53, pp. 65–78, 2003.
- [8] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2508–2530, June 2006.
- [9] S. Kar and J. M. F. Moura, "Distributed consensus algorithms in sensor networks with imperfect communication: Link failures and channel noise," *IEEE Transactions on Signal Processing*, vol. 57, no. 1, pp. 355–369, Jan. 2009.
- [10] S. Kar and J. M. F. Moura, "Convergence rate analysis of distributed gossip (linear parameter) estimation: Fundamental limits and tradeoffs," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 4, pp. 674–690, Aug. 2011.
- [11] A. Nedić, A. Olshevsky, A. Ozdaglar, and J. N. Tsitsiklis, "On distributed averaging algorithms and quantization effects," *IEEE Transactions on Automatic Control*, vol. 54, no. 11, pp. 2506–2517, Nov. 2009.
- [12] S. Y. Tu and A. H. Sayed, "Diffusion strategies outperform consensus strategies for distributed estimation over adaptive networks," *IEEE Transactions on Signal Processing*, vol. 60, no. 12, pp. 6217–6234, Dec. 2012.
- [13] F. S. Cattivelli and A. H. Sayed, "Diffusion lms strategies for distributed estimation," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1035–1048, March 2010.
- [14] R. Olfati-Saber and R. M. Murray, "Consensus problems in networks of agents with switching topology and time-delays," *IEEE Transactions on Automatic Control*, vol. 49, no. 9, pp. 1520–1533, Sept. 2004.
- [15] A. Nedić and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Transactions on Automatic Control*, vol. 54, no. 1, pp. 48–61, Jan. 2009.
- [16] D. Blatt, A. O. Hero, and H. Gauchman, "A convergent incremental gradient method with a constant step size," *SIAM Journal on Optimization*, vol. 18, no. 1, pp. 29–51, 2007.
- [17] I. Lobel and A. Ozdaglar, "Distributed subgradient methods for convex optimization over random networks," *IEEE Transactions on Automatic Control*, vol. 56, no. 6, pp. 1291–1306, June 2011.
- [18] J. C. Duchi, A. Agarwal, and M. J. Wainwright, "Dual averaging for distributed optimization: Convergence analysis and network scaling," *IEEE Transactions on Automatic Control*, vol. 57, no. 3, pp. 592–606, March 2012.
- [19] K. I. Tsianos and M. G. Rabbat, "Distributed dual averaging for convex optimization under communication delays," in *Proc. 2012 American Control Conference (ACC)*, June 2012, pp. 1067–1072.
- [20] K. I. Tsianos, S. Lawlor, and M. G. Rabbat, "Push-sum distributed dual averaging for convex optimization," in *Proc. 51st IEEE Conference on Decision and Control (CDC)*, Dec. 2012, pp. 5453–5458.
- [21] D. Yuan, S. Xu, H. Zhao, and L. Rong, "Distributed dual averaging method for multi-agent optimization with quantized communication," *Systems & Control Letters*, vol. 61, no. 11, pp. 1053 – 1061, 2012.
- [22] S. Lee, A. Nedić, and M. Raginsky, "Coordinate dual averaging for decentralized online optimization with nonseparable global objectives," *IEEE Transactions on Control of Network Systems*, vol. PP, no. 99, pp. 1–1, 2016.
- [23] S. Hosseini, A. Chapman, and M. Mesbahi, "Online distributed convex optimization on dynamic networks," *IEEE Transactions on Automatic Control*, vol. 61, no. 11, pp. 3545–3550, Nov 2016.
- [24] F. R. K. Chung, *Spectral Graph Theory*, American Mathematical Society, Dec. 1996.
- [25] B. Smith, P. Bjorstad, and W. Gropp, *Domain decomposition: parallel multilevel methods for elliptic partial differential equations*, Cambridge university press, 2004.
- [26] A. Ghosh and S. Boyd, "Growing well-connected graphs," in *Proceedings of the 45th IEEE Conference on Decision and Control*, Dec. 2006, pp. 6605–6611.
- [27] S. Boyd, "Convex optimization of graph Laplacian eigenvalues," in *International Congress of Mathematicians*, 2006, pp. 1311–1319.
- [28] D. Xue, A. Gusrialdi, and S. Hirche, "A distributed strategy for near-optimal network topology design," in *Proc. 21th International Symposium on Mathematical Theory of Networks and Systems*, 2014, pp. 7–14.
- [29] P.-Y. Chen and A. O. Hero, "Assessing and safeguarding network resilience to nodal attacks," *IEEE Communications Magazine*, vol. 52, no. 11, pp. 138–143, Nov. 2014.
- [30] David A. Levin, Yuval Peres, and Elizabeth L. Wilmer, *Markov chains and mixing times*, American Mathematical Society, 2006.
- [31] S. Liu, P.-Y. Chen, and A. O. Hero, "Distributed optimization for evolving networks of growing connectivity," to appear in *ICASSP'17*, https://www.researchgate.net/publication/308305484_Distributed_Optimization_for_Evolving_Networks_of_Growing_Connectivity.
- [32] M. Penrose, *Random Geometric Graphs*, Oxford University Press, New York, 2003.