# Approximate Projection Methods for Decentralized Optimization with Functional Constraints

Soomin Lee, *Member, IEEE,* and Michael M. Zavlanos, *Member, IEEE*

*Abstract*—We consider distributed convex optimization problems that involve a separable objective function and nontrivial functional constraints, such as Linear Matrix Inequalities (LMIs). We propose a decentralized and computationally inexpensive algorithm which is based on the concept of approximate projections. Our algorithm is one of the consensus based methods in that, at every iteration, each agent performs a consensus update of its decision variables followed by an optimization step of its local objective function and local constraints. Unlike other methods, the last step of our method is not an Euclidean projection onto the feasible set, but instead a subgradient step in the direction that minimizes the local constraint violation. We propose two different averaging schemes to mitigate the disagreements among the agents' local estimates. We show that the algorithms converge almost surely, i.e., every agent agrees on the same optimal solution, under the assumption that the objective functions and constraint functions are nondifferentiable and their subgradients are bounded. We provide simulation results on a decentralized optimal gossip averaging problem, which involves SDP constraints, to complement our theoretical results.

## I. INTRODUCTION

**D**ECENTRALIZED optimization has been extensively studied in recent years due to a variety of applications in machine learning, signal processing, and control for robotic networks, sensor networks, power networks, and wireless communication networks [1]–[5]. A number of problems arising in these areas can be cast as distributed convex optimization problems over multiagent networks, where individual agents cooperatively try to minimize a common cost function over a common constraint set in the absence of full knowledge about the global problem structure. The main feature of carrying these optimizations over networks is that the agents can only communicate with their neighboring agents. This communication structure can be cast as a graph, often directed and/or time-varying.

The literature on distributed optimization methods is vast and involves first-order methods in the primal domain, the dual domain, augmented Lagrangian methods, or Newton methods, to name a few. Here we discuss methods that are most closely related to the method under consideration. Among those, one of the most well-studied techniques are the so called *consensus-based* optimization algorithms [6]–[23] (see also the literature for the consensus problem itself [4], [5], [24]),

where the goal is to repeatedly average the estimates of all agents in a decentralized fashion in order to obtain a network-wide consensus. Between the averaging steps, each agent usually performs a single local optimization step. Overall, the agents use their local information to cooperatively steer the consensus point toward the optimal set of the global problem.

Decentralized algorithms that fall in this class of methods can be distinguished based on which averaging scheme or optimization method is used, and in which space (primal or dual) the iterates are maintained. All these algorithms often require expensive optimization steps or exact projections on a complicated constraint set at every iteration. Such intensive computations, however, require time and may shorten the lifespan of certain systems, such as wireless sensor networks or robotic networks.

In this work, we propose a new approximate projection based decentralized algorithm and prove its convergence. Our work in this paper is an extension of the author's previous work [25]. Specifically, we use the same local information exchange model and gradient descent algorithm as in [25], but a different projection method motivated by the work in [26]. In contrast to [25], our contribution can be summarized as follows: (1) Instead of using the Euclidean projection, we approximate it by measuring the constraint violation and taking a subgradient step minimizing this violation; (2) We show convergence under milder assumptions. Specifically, we remove the smoothness assumption in the objective functions; (3) We propose two different averaging schemes to mitigate the disagreements among the agents' local estimates, one of which can lift the doubly stochasticity assumption on the weight matrices.

Considering that projections have a closed form solution only in a few special cases of constraints, our new algorithm is more general and can be applied to a wider class of problems including Semidefinite Programming (SDP), where the constraints are represented by Linear Matrix Inequalities (LMIs). It is well known that even finding a feasible point that satisfies a handful of LMIs is a difficult problem on its own. The work in this paper is also related to the centralized random projection algorithms for convex constrained optimization [27] and convex feasibility problems [28]. Other related works are [29]–[31], where optimization problems with uncertain constraints have been considered by finding probabilistic feasible solutions through random sampling of constraints.

The paper is organized as follows. In Section II, we formulate the optimization problem under consideration and discuss specific problems of interest. In Section III, we provide our decentralized algorithm based on random approximate projections, discuss the communication scheme employed by the agents, state assumptions and the main results of this paper.

In Section IV, we first review some necessary results and lemmas from existing literature, provide proofs of required lemmas, and then present the proofs of the main results discussed in Section III. In Section V, we present simulation results for a decentralized SDP problem, which is optimal decentralized gossip averaging. We conclude the paper with some comments in Section VI.

**Notation:** All vectors are viewed as column vectors. We write $x^\top$ to denote the transpose of a vector $x$. The scalar product of two vectors $x$ and $y$ is $\langle x, y \rangle$. For vectors associated with agent $i$ at time $k$, we use subscripts $i, k$ such as, for example, $p_{i,k}$, $x_{i,k}$, etc. Unless otherwise stated, $\|\cdot\|$ represents the standard Euclidean norm. For a set $S$, we use $|S|$ to denote its cardinality. For a matrix $A \in \mathbb{R}^{n \times n}$, we use $[A]_{ij}$ to denote the entry of the $i$-th row and $j$-th column and $\|A\|_F$ to denote the Frobenius norm $\|A\|_F = \left( \sum_{i,j=1}^{n} ([A]_{ij})^2 \right)^{1/2}$. We use $\mathsf{Tr}A$ to denote the trace of $A$, i.e., $\mathsf{Tr}A = \sum_{i=1}^{n} [A]_{ii}$. We denote by $\mathbb{S}_m$ and $\mathbb{S}_m^+$ the space of $m \times m$ real symmetric and real symmetric positive semidefinite matrices, respectively. The matrix inequality $A \preceq 0$ means $-A$ is positive semidefinite. We use $\mathbf{1}$ and $\mathbf{0}$ to denote vectors of all ones and zeros. The identity matrix is denoted by $I$. We use $\mathsf{Pr}\{Z\}$ and $\mathsf{E}[Z]$ to denote the probability and the expectation of a random variable $Z$. We write $\mathrm{dist}(x, \mathcal{X})$ for the distance of a vector $x$ from a closed convex set $\mathcal{X}$, i.e., $\mathrm{dist}(x, \mathcal{X}) = \min_{v \in \mathcal{X}} \|v - x\|$. We use $\Pi_{\mathcal{X}}[x]$ for the Euclidean projection of a vector $x$ on the set $\mathcal{X}$, i.e., $\Pi_{\mathcal{X}}[x] = \arg\min_{v \in \mathcal{X}} \|v - x\|^2$. We often abbreviate *almost surely* and *independent identically distributed* as *a.s.* and *i.i.d.*, respectively.

## II. PROBLEM DEFINITION

Consider a multiagent network system whose communication at time $k$ is governed by a digraph $\mathcal{G}_k = (\mathcal{V}, \mathcal{E}_k)$, where $\mathcal{V} = \{1, \ldots, N\}$ and $\mathcal{E}_k \subseteq \mathcal{V} \times \mathcal{V}$. If there exists a directed link from agent $j$ to $i$, which we denote by $(j, i)$, agent $j$ may send its information to agent $i$. Thus, each agent $i \in \mathcal{V}$ can directly receive information only from the agents in its in-neighborhood

$$\mathcal{N}_{i,k}^{\mathrm{in}} = \{j \in \mathcal{V} \mid (j, i) \in \mathcal{E}_k\} \cup \{i\}, \tag{1}$$

and send information only to the agents in its out-neighborhood

$$\mathcal{N}_{i,k}^{\mathrm{out}} = \{j \in \mathcal{V} \mid (i, j) \in \mathcal{E}_k\} \cup \{i\}, \tag{2}$$

where in both $\mathcal{N}_{i,k}^{\mathrm{in}}$ and $\mathcal{N}_{i,k}^{\mathrm{out}}$, we assume there exists a self-loop $(i, i)$ for all $i \in \mathcal{V}$. Also, we use $\deg_{i,k}$ to denote the in-degree of node $i$ at time $k$, i.e.,

$$\deg_{i,k} = |\mathcal{N}_{i,k}^{\mathrm{in}}|. \tag{3}$$

### A. Problem Statement

Our goal is to let the network of agents cooperatively solve the following convex minimization problem:

$$\min_{x} \; f(x) \triangleq \sum_{i \in \mathcal{V}} f_i(x) \tag{4}$$

$$\text{s.t. } x \in \mathcal{X}, \quad \mathcal{X} \triangleq \mathcal{X}_0 \bigcap \left( \bigcap_{i \in \mathcal{V}} \mathcal{X}_i \right),$$

where only agent $i$ knows the function $f_i : \mathbb{R}^n \to \mathbb{R}$ and the constraint set $\mathcal{X}_i \subseteq \mathbb{R}^n$. The set $\mathcal{X}_0 \subseteq \mathbb{R}^n$ is common to all agents and assumed to have some simple structure in the sense that the projection onto $\mathcal{X}_0$ can be made easily (e.g., a box, ball, probability simplex, or even $\mathbb{R}^n$). Note that the common constraint set, i.e., $\mathcal{X}_0 = \mathcal{X}_i$ for all $i \in \mathcal{V}$, is a special case of this problem definition. We assume that the set of optimal solutions $\mathcal{X}^* = \arg\min_{x \in \mathcal{X}} f(x)$ is nonempty.

We assume each agent $i$'s local constraint set $\mathcal{X}_i$ consists of one or more algebraic inequalities, which we denote by

$$\mathcal{X}_i = \{x \in \mathbb{R}^n \mid g(x, \omega) \leq 0, \ \forall \omega \in \Omega_i\},$$

where $\Omega_i$ is a finite collection of indices. From this definition, the feasible set $\mathcal{X}$ can be precisely represented as

$$\mathcal{X} = \{x \in \mathcal{X}_0 \mid g(x, \omega) \leq 0, \ \forall \omega \in \Omega_i, i \in \mathcal{V}\}.$$

Note that some of the inequalities may overlap across different agents, i.e., $\Omega_i \cap \Omega_j$ for $i \neq j$ can be either empty or nonempty.

We also consider an equivalent epigraph form of problem (4) by introducing a new set of variables $t = [t_1 \ \ldots \ t_N]^\top \in \mathbb{R}^N$. Consider that the local constraint set $\mathcal{X}_i$ for $i \in \mathcal{V}$ now includes the additional inequality constraint $f_i(y) \leq t_i$. Then, problem (4) is equivalent to:

$$\min_{x} \; a^\top x \tag{5}$$

$$\text{s.t. } y \in \mathcal{X}, \quad \mathcal{X} \triangleq \mathcal{X}_0 \bigcap \left( \bigcap_{i \in \mathcal{V}} \mathcal{X}_i \right),$$

where $x = [y^\top \ t^\top]^\top$ and $a = [\mathbf{0}^\top \ \mathbf{1}^\top]^\top$. Note that this reformulation has also been introduced in [32] for solving distributed robust convex optimization. In the case of a single agent, the algorithm can be also seen as scenario approach for solving robust convex optimization (see e.g., [33]–[35] and references therein).

### B. Problems of Interest

Problems of particular interest are those involving lots of nontrivial constraints on which exact projections are impossible or computationally intractable. Here we provide two such examples:

1) Robust Linear Inequalities:

$$\mathcal{X}_i = \Big\{x \in \mathbb{R}^n \mid A(\omega)x \leq b(\omega), \ \forall \omega \text{ such that}$$

$$\|A(\omega) - A_0\|_{\mathrm{op}} \leq r_1 \text{ and } \|b(\omega) - b_0\|_{\mathrm{op}} \leq r_2 \Big\},$$

where $A_0 \in \mathbb{R}^{m \times n}, b_0 \in \mathbb{R}^m$ are nominal data, $r_1, r_2 \geq 0$ are the levels of uncertainty, and $\|\cdot\|_{\mathrm{op}}$ denotes an operator norm. Here we can not handle each row of $A(\omega)x \leq b(\omega)$ separately as in [25] due to the matrix operator norm $\|\cdot\|_{\mathrm{op}}$.

2) Linear Matrix Inequalities:

$$\mathcal{X}_i = \Big\{x \in \mathbb{R}^n \mid A_0(\omega) + \sum_{j=1}^{n} x_j A_j(\omega) \preceq 0, \ \forall \omega \in \Omega_i \Big\}, \tag{6}$$

where $A_j(\omega) \in \mathbb{S}_m$ for $j = 0, 1, \ldots, n$, $\omega \in \Omega_i$ are given matrices. The inequalities in (6) are referred to as *linear matrix inequalities* (LMIs). A semidefinite

**Algorithm 1** Decentralized Approximate Projection (DAP)

---

Let $x_{i,0} \in \mathcal{X}_0$ for $i \in \mathcal{V}$ and the nonnegative parameter $\{\alpha_k\}_{k \geq 0}$ be given.
Set $k := 0$
**while** Maximum iteration number is reached **do**
  Each agent $i$ updates $x_{i,k}$ according to

$$p_{i,k} = \sum_{j \in \mathcal{V}} [W_k]_{ij} x_{j,k-1} \qquad (7a)$$

$$v_{i,k} = \Pi_{\mathcal{X}_0} [p_{i,k} - \alpha_k s_{i,k}] \qquad (7b)$$

$$x_{i,k} = \Pi_{\mathcal{X}_0} \left[ v_{i,k} - \frac{g^+(v_{i,k}, \omega_{i,k})}{\|d_{i,k}\|^2} d_{i,k} \right] \qquad (7c)$$

  Set $k := k + 1$
**end while**

---

programming (SDP) problem has one or more LMI constraints. Finding a feasible point of the set (6) is often a difficult problem on its own.

Note that the inequalities in (6) can represent a wide variety of convex constraints (see [36] for more details). For example, quadratic inequalities, inequalities involving matrix norms, and various inequality constraints arising in robust control such as Lyapunov and quadratic matrix inequalities can be all cast as LMIs in (6). When all matrices $A_j(\omega)$ in (6) are diagonal, the LMIs reduce to regular linear inequalities.

## III. ALGORITHM, ASSUMPTIONS, AND MAIN RESULTS

Our goal is to design a decentralized protocol by which each agent $i \in \mathcal{V}$ maintains a sequence of the local copy $\{x_{i,k}\}_{k \geq 0}$ converging to the same point in $\mathcal{X}^*$ as $k$ goes to infinity. Since we assume that the local constraint sets $\mathcal{X}_i$'s are nontrivial, we do not find an exact projection onto $\mathcal{X}_i$ at each step of the algorithm. Instead, at iteration $k$, each agent $i$ randomly generates an index $\omega_{i,k} \in \Omega_i$ and makes an *approximate projection* on the selected inequality $g(\cdot, \omega_{i,k}) \leq 0$.

### A. Decentralized Algorithm with Approximate Projections

We formally present our decentralized algorithm, named the Decentralized Approximate Projection (DAP), in Algorithm 1. Each agent $i$ maintains a sequence $\{x_{i,k}\}_{k \geq 0}$. The element $x_{i,k}$ of the sequence can be seen as the agent $i$'s estimate of the decision variable $x$ at time $k$. Let $g^+(x, \omega)$ denote the function that measures the violation of the constraint $g(\cdot, \omega)$ at $x$, i.e., $g^+(x, \omega) \triangleq \max\{g(x, \omega), 0\}$.

At $k = 0$, the estimates $x_{i,0}$ are locally initialized such that $x_{i,0} \in \mathcal{X}_0$. At time step $k$, all agents $j \in \mathcal{V}$ broadcast their previous estimates $x_{j,k-1}$ to all of the nodes in their out-neighborhood, i.e., to all agents $i$ such that $(i, j) \in \mathcal{E}_k$. Then, each agent $i \in \mathcal{V}$ updates $x_{i,k}$ using (7a)-(7c), where $W_k$ is a nonnegative $N \times N$ weight matrix, $\{\alpha_k\}$ is a positive sequence of nonincreasing stepsizes; $s_{i,k}$ is a subgradient of the function $f_i$ at $p_{i,k}$; $\omega_{i,k}$ is a random variable taking values in the index set $\Omega_i$; and $d_{i,k}$ is a subgradient of $g^+(\cdot, \omega_{i,k})$ evaluated at $v_{i,k}$. The vector $d_{i,k}$ is chosen such that $d_{i,k} \in \partial g^+(v_{i,k}, \omega_{i,k})$ if $g^+(v_{i,k}, \omega_{i,k}) > 0$, and $d_{i,k} = d_i$ for some $d_i \neq 0$, if $g^+(v_{i,k}, \omega_{i,k}) = 0$.

More specifically, in (7a), each agent $i$ calculates a weighted average of the received messages (including its own message $x_{i,k-1}$) to obtain $p_{i,k}$. Specifically, $[W_k]_{ij} \geq 0$ is the weight that agent $i$ allocates to the message $x_{j,k-1}$. This communication step is decentralized since the weight matrix $W_k$ respects the topology of the graph $\mathcal{G}_k$, i.e., $[W_k]_{ij} > 0$ only if $(j, i) \in \mathcal{E}_k$ and $[W_k]_{ij} = 0$, otherwise. In (7b), each agent $i$ adjusts the average $p_{i,k}$ in the direction of the negative subgradient $(-s_{i,k})$ of its local objective $f_i$ to obtain $v_{i,k}$. The adjusted average is projected back to the simple set $\mathcal{X}_0$. In (7c), agent $i$ observes a random realization of $\omega_{i,k} \in \Omega_i$ and measures the feasibility violation of the selected component constraint $g(\cdot, \omega_{i,k})$ at $v_{i,k}$. If $g^+(v_{i,k}, \omega_{i,k}) > 0$, it calculates a subgradient $d_{i,k} \in \partial g^+(v_{i,k}, \omega_{i,k})$ and takes an additional subgradient step with the stepsize $\frac{g^+(v_{i,k}, \omega_{i,k})}{\|d_{i,k}\|^2}$ to minimize this violation. If $g^+(v_{i,k}, \omega_{i,k}) = 0$, then the current point $v_{i,k}$ already satisfies the selected inequality $g(\cdot, \omega_{i,k}) \leq 0$. In this case, there is no need to move the point further into the selected set. Therefore, the approximate projection step (7c) is just omitted.

Note that the description of the DAP algorithm is only conceptual at this moment since we have not specified the parameters $\{\alpha_k\}$ and $\{W_k\}$ yet. The stepsizes $\{\alpha_k\}$ should be nonnegative, nonincreasing and such that

$$\sum_{k=1}^{\infty} \alpha_k = \infty \text{ and } \sum_{k=1}^{\infty} \alpha_k^2 < \infty. \qquad (8)$$

For the sequence of weight matrices $\{W_k\}$, we assume the following.

*Assumption 1: For all $k \geq 1$,*
(a) $[W_k]_{ij} \geq 0$ *for all $i, j \in \mathcal{V}$ and $[W_k]_{ij} = 0$ only if $j \notin \mathcal{N}_{i,k}^{in}$.*
(b) *There exists a scalar $\nu \in (0, 1)$ such that $[W_k]_{ij} \geq \nu$ only if $j \in \mathcal{N}_{i,k}^{in}$.*
(c) $\sum_{j \in \mathcal{V}} [W_k]_{ij} = 1$ *for all $i \in \mathcal{V}$, and $\sum_{i \in \mathcal{V}} [W_k]_{ij} = 1$ for all $j \in \mathcal{V}$.*

Condition (a) ensures that the weight matrices $W_k$ respect the underlying topology $\mathcal{G}_k$ for every $k$ so that the communication is indeed decentralized. The lower boundedness of the weights in (b) is required to show consensus among all agents (see [37] for more details) but the agents need not know the $\nu$ value in running the algorithm. Condition (a) and (c) imply doubly stochasticity of the matrices $W_k$.

Note that assuming doubly stochasticity of $W_k$ for all $k \geq 1$ often requires a global view of the network (unless the underlying graph is regular or fully connected) and not all directed graphs admit a doubly stochastic matrix [38], [39]. The linear objective function in the epigraph formulation (5) allows us to lift this assumption and to use a weight-imbalanced mixing matrix. For example, we can assume a time-invariant network, i.e., $\mathcal{G}_k = \mathcal{G}$ for $k \geq 1$, and employ a row stochastic matrix $W$ whose entries are defined as follows:

$$[W]_{ij} = \begin{cases} \frac{1}{\deg_i} & \text{if } j \in \mathcal{N}_i^{in}, \\ 0 & \text{otherwise.} \end{cases} \qquad (9)$$

Note that this choice of weights also respects the underlying topology $\mathcal{G}$. Moreover, the matrix $W$ is not necessarily column stochastic.

Recall that our problems of interest involve a large number of constraints. Therefore, the random selection of a constraint in (7c) serves as a computationally efficient alternative to finding the most violated constraint, which typically has significantly higher per-iteration complexity. Another situation that necessitates the random selection approach is when the constraints are not fully given in advance, but are rather revealed in a sequential fashion (as in online optimization).

Note that the step (7c) guarantees that $x_{i,k} \in \mathcal{X}_0$ for all $k \geq 0$ and $i \in \mathcal{V}$, but it does not necessarily guarantee $x_{i,k} \in \mathcal{X}$. Nevertheless, in Section IV, we show that $x_{i,k}$ for all $i \in \mathcal{V}$ asymptotically achieve feasibility, i.e., $\lim_{k\to\infty} \|x_{i,k} - \Pi_{\mathcal{X}}[p_{i,k}]\| = 0$ for all $i \in \mathcal{V}$.

To further explain the step (7c), let us consider the two particular cases mentioned in Section II-B.

1) Let $c^+$ denote a projection of a vector $c \in \mathbb{R}^m$ onto the nonnegative orthant. We introduce a scalar function in order to handle all the rows of the inequality $A(\omega)x \leq b$ concurrently,

$$g^+(x,\omega) = \|(A(\omega)x - b)^+\|$$

which is convex in $x$ for any given $\omega \in \Omega_i$ and $i \in \mathcal{V}$ [40, Chapter 3.2]. Then, it is straightforward to see that its subgradient can be calculated as

$$\partial g^+(x,\omega) = \frac{A(\omega)^\top (A(\omega)x - b)^+}{\|(A(\omega)x - b)^+\|},$$

if $g^+(x,\omega) > 0$, and we can use $\partial g^+(x,\omega) = 0$, otherwise.

2) Let us define the projection $A^+$ of a real symmetric matrix $A$ onto the cone of positive semidefinite matrices. For any $A \in \mathbb{S}^m$, we can find an eigenvalue decomposition $A = B\Lambda B^\top$, where $B$ is an orthogonal matrix and $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_m)$. Then, its projection is given by

$$A^+ = B\Lambda^+ B^\top,$$

where $\Lambda^+ = \mathrm{diag}(\lambda_1^+, \ldots, \lambda_m^+)$ with $\lambda_i^+ = \max\{0, \lambda_i\}$ [41]. Let us define

$$A(x,\omega) \triangleq A_0(\omega) + \sum_{j=1}^n x_j A_j(\omega).$$

Then, the amount of violation of the corresponding LMI constraint $A(x,\omega) \preceq 0$ can be measured by the following convex scalar function:

$$g^+(x,\omega) = \|A^+(x,\omega)\|_F. \qquad (10)$$

By direct calculations, it is not difficult to see that its subgradient is given by

$$\partial g^+(x,\omega) = \frac{1}{g^+(x,\omega)} \begin{pmatrix} \mathrm{Tr} A_1 A^+(x,\omega) \\ \vdots \\ \mathrm{Tr} A_n A^+(x,\omega) \end{pmatrix}, \quad (11)$$

if $g^+(x,\omega) > 0$, and we can use $\partial g^+(x,\omega) = 0$, otherwise [26, Lemma 1].

**Remark** Note that the computational complexity of step (7c) depends on the type of the function $g(\cdot, \omega_{i,k})$. If $g(\cdot, \omega_{i,k})$ is a general convex function, it takes $O(1)$ computations for the evaluation of $g^+(\cdot, \omega_{i,k})$ and $O(n)$ computations for the evaluation of the gradient $d_{i,k}$. If $g(\cdot, \omega_{i,k})$ is an LMI constraint, it takes $O(m^3)$ computations in the worst-case for the eigenvalue decomposition and $O(m^2)$ for the computation of the Frobenius norm (cf. Eq. (10)). We would also need $O(m^2 n)$ computations for computing the traces (cf. Eq. (11)). This eigenvalue decomposition is necessary for projection (or approximate projection) onto the cone of positive semidefinite matrices. ∎

It is also worth mentioning that the algorithm (7a)-(7c) includes the method that has been proposed in [25] as a special case. In order to see this, let $\mathcal{X}_0 = \mathbb{R}^n$ and $g(x,\omega_{i,k}) = \mathrm{dist}(x, \mathcal{X}_i^{\omega_{i,k}})$, where $\mathcal{X}_i^{\omega_{i,k}} = \{x \in \mathbb{R}^n \mid g(x,\omega_{i,k}) \leq 0\}$. Then, it is not difficult to see that

$$d_{i,k} = \frac{v_{i,k} - \Pi_{\mathcal{X}_i^{\omega_{i,k}}}[v_{i,k}]}{\|v_{i,k} - \Pi_{\mathcal{X}_i^{\omega_{i,k}}}[v_{i,k}]\|},$$

and since $\mathrm{dist}(v_{i,k}, \Pi_{\mathcal{X}_i^{\omega_{i,k}}}) = \|v_{i,k} - \Pi_{\mathcal{X}_i^{\omega_{i,k}}}[v_{i,k}]\|$, the steps (7b)-(7c) reduce to

$$x_{i,k} = \Pi_{\mathcal{X}_i^{\omega_{i,k}}}[p_{i,k} - \alpha_k s_{i,k}],$$

which is exactly the algorithm in [25].

### B. Assumptions

For the optimization problem (4), we make the following assumptions on the set $\mathcal{X}_0$, the objective functions $f_i(x)$ for $i \in \mathcal{V}$, and the constraint functions $g(x,\omega)$ for $\omega \in \Omega_i$ and $i \in \mathcal{V}$.

*Assumption 2:*
(a) *The set $\mathcal{X}_0$ is nonempty, closed and convex.*
(b) *The function $f_i(x)$, for each $i \in \mathcal{V}$, is defined and convex (not necessarily differentiable) over some open set that contains $\mathcal{X}_0$.*
(c) *The subgradients $s \in \partial f_i(x)$ are uniformly bounded over the set $\mathcal{X}_0$. That is, for all $i \in \mathcal{V}$, there is a scalar $C_{f_i}$ such that for all $s \in \partial f_i(x)$ and $x \in \mathcal{X}_0$,*

$$\|s\| \leq C_{f_i}.$$

(d) *The function $g(x,\omega)$, for each $\omega \in \Omega_i$ and $i \in \mathcal{V}$, is defined and convex in $x$ (not necessarily differentiable) over some open set that contains $\mathcal{X}_0$.*
(e) *The subgradients $d \in \partial g^+(x,\omega)$ are uniformly bounded over the set $\mathcal{X}_0$. That is, there is a scalar $C_g$ such that for all $d \in \partial g^+(x,\omega)$, $x \in \mathcal{X}_0$, $\omega \in \Omega_i$, and $i \in \mathcal{V}$,*

$$\|d\| \leq C_g.$$

By Assumption 2, the subdifferentials $\partial f_i(x)$ and $\partial g^+(x,\omega)$ are nonempty over $\mathcal{X}_0$. It also implies that for any $i \in \mathcal{V}$ and $x, y \in \mathcal{X}_0$,

$$|f_i(x) - f_i(y)| \leq C_{f_i} \|x - y\|, \qquad (12)$$

and for any $\omega \in \Omega_i$, $i \in \mathcal{V}$, and $x, y \in \mathcal{X}_0$,

$$|g^+(x,\omega) - g^+(y,\omega)| \leq C_g \|x - y\|. \qquad (13)$$

One sufficient condition for Assumption 2(c) and 2(e) is that the set $\mathcal{X}_0$ is compact.

We also require the following two assumptions.

*Assumption 3:* We assume that $\omega_{i,k} \in \Omega_i$ are i.i.d. samples from some probability distribution on $\Omega_i$ and independent across agents. Furthermore, each $\Omega_i$ is a finite set and each element of $\Omega_i$ is generated with nonzero probability, i.e., for any $\omega \in \Omega_i$ and $i \in \mathcal{V}$

$$\mathsf{Pr}\{\omega \mid \omega \in \Omega_i\} > 0$$

Let $\mathcal{F}_k$ denote the algorithm's history up to time $k$. i.e.,

$$\mathcal{F}_k = \{x_{i,0}, \; (\omega_{i,t}, \; 1 \le t \le k), \; i \in \mathcal{V}\},$$

and $\mathcal{F}_= \{x_{i,0}, \; i \in \mathcal{V}\}$.

*Assumption 4:* For all $i \in \mathcal{V}$, there exists a constant $c > 0$ such that for all $x \in \mathcal{X}_0$

$$\text{dist}^2(x, \mathcal{X}) \le c \mathsf{E}\left[(g^+(x, \omega))^2 \mid \mathcal{F}_{k-1}\right],$$

*where the expectation is taken with respect to the set $\Omega_i$.*
The upper bound in Assumption 4 is known as *global error bound* and is crucial for the convergence analysis of our method (7a)-(7c). Sufficient conditions for this bound have been shown in [42] and [43], which require the existence of a Slater point, i.e., let $\mathcal{X}_0 = \{x \mid g_0(x) \le 0\}$, then there exists a point $\tilde{x}$ such that $g_0(\tilde{x}) < 0$ and $g(\tilde{x}, \omega) < 0$ for all $\omega$. When each function $g(\cdot, \omega)$ and $g_0(\cdot)$ is either a linear equality or inequality, Assumption 4 is called *linear regularity* and can be shown to hold by using the results in [44] and [45] (see also [46]–[48]).

The inter-agent communication relies on the time-varying graph sequence $\mathcal{G}_k = (\mathcal{V}, \mathcal{E}_k)$, for $k \ge 1$. A key assumptions on these communication graphs is the following:

*Assumption 5:* There exists a scalar $Q$ such that the graphs $\left(\mathcal{V}, \bigcup_{\ell=0,\ldots,Q-1} \mathcal{E}_{k+\ell}\right)$ are strongly connected for all $k \ge 1$.

Assumption 5 ensures that there exists a path from one agent to every other agent within any bounded interval of length $Q$. We say that such a sequence of graphs is $Q$-strongly connected.

### C. Main Results

We now provide two convergence results for DAP, Proposition 1 for problem (4) and Proposition 2 for problem (5), respectively, for which we use different assumptions on the network. The first proposition states a convergence result which holds under a $Q$-strongly connected time-varying sequence of graphs $\{\mathcal{G}_k\}$ and the corresponding doubly stochastic matrices $\{W_k\}$ which respect the graph topologies.

*Proposition 1:* Let Assumptions 1 - 5 hold and the optimal set $\mathcal{X}^*$ of (4) be nonempty. Let the nonnegative and nonincreasing stepsize $\{\alpha_k\}$ satisfy conditions in (8). Then, the iterates $\{x_{i,k}\}$ generated by each agent $i \in \mathcal{V}$ via DAP in Algorithm 1 converge almost surely to the same point in the optimal set $\mathcal{X}^*$ of (4), i.e., for a random point $x^* \in \mathcal{X}^*$

$$\lim_{k\to\infty} x_{i,k} = x^* \quad \text{for all } i \in \mathcal{V} \quad a.s.$$

The second proposition states a convergence result which holds under a connected time-invariant graph $\mathcal{G}$ and the corresponding row stochastic matrix $W$ which respects the graph

topology. Note that for this case we can set $f_i(x) = \frac{1}{|\mathcal{V}|}a^\top x$ for all $i \in \mathcal{V}$ and $s_{i,k} = a$ for all $i \in \mathcal{V}$ and $k \ge 0$.

*Proposition 2:* Let Assumptions 2 - 4 hold and the optimal set $\mathcal{X}^*$ of (5) be nonempty. Let the nonnegative and nonincreasing stepsize $\{\alpha_k\}$ satisfy conditions in (8). Let $W_k = W$ for all $k \ge 1$ and Assumption 5 hold with $Q = 1$. Then, the iterates $\{x_{i,k}\}$ generated by each agent $i \in \mathcal{V}$ via DAP in Algorithm 1 with the choice of weight in (9) converge almost surely to the same point in the optimal set $\mathcal{X}^*$ of (5), i.e., for a random point $x^* \in \mathcal{X}^*$

$$\lim_{k\to\infty} x_{i,k} = x^* \quad \text{for all } i \in \mathcal{V} \quad a.s.$$

In the next paragraph, we provide some intuition on why DAP still converges even with a choice of a row stochastic matrix $W$. Note that the use of row stochastic matrix results in "biased" consensus, which is related to the left-eigenvector, see e.g., [49]. The following lemma states this well-known result.

*Lemma 1:* Let Assumption 5 hold with $Q = 1$ and $W = W_k$ for any $k \ge 1$. Then, there exists a normalized left-eigenvector $\pi \in \mathbb{R}^N$ such that

$$\pi^\top W = \pi^\top.$$

Moreover, $[\pi]_i > 0$ for all $i \in \mathcal{V}$.

This will lead the algorithm to converge to an optimal solution of a biased objective function, $\sum_{i\in\mathcal{V}}[\pi]_i f_i(x)$, instead of the true objective $\sum_{i\in\mathcal{V}} f_i(x)$. However, since the $f_i(x)$'s in the reformulated epigraph form in (5) is now linear, the biased objective function is just the same as the original function, i.e., $\sum_{i\in\mathcal{V}}[\pi]_i f_i(x) = \sum_{i\in\mathcal{V}}[\pi]_i a^\top x = a^\top x$.

## IV. CONVERGENCE ANALYSIS

In this section, we are concerned with demonstrating the convergence results stated in Proposition 1 and 2. First we review some lemmas from existing literature that are necessary in our analysis.

### A. Preliminary Results

First we state a non-expansiveness property of the projection operator (see [50] for its proof).

*Lemma 2:* Let $\mathcal{X} \subseteq \mathbb{R}^n$ be a nonempty closed convex set. The function $\Pi_{\mathcal{X}} : \mathbb{R}^n \to \mathcal{X}$ is nonexpansive, i.e.,

$$\|\Pi_{\mathcal{X}}[x] - \Pi_{\mathcal{X}}[y]\| \le \|x - y\| \quad \text{for all } x, y \in \mathbb{R}^n.$$

In our analysis of the algorithm, we also make use of the following convergence result due to Robbins and Siegmund (see [51, Lemma 10-11, p. 49-50]).

*Theorem 1:* Let $\{v_k\}$, $\{u_k\}$, $\{a_k\}$ and $\{b_k\}$ be sequences of non-negative random variables such that

$$\mathsf{E}[v_{k+1}|\mathcal{F}_k] \le (1+a_k)v_k - u_k + b_k \quad \text{for all } k \ge 0 \quad a.s.$$

*where $\mathcal{F}_k$ denotes the collection $v_0, \ldots, v_k, \; u_0, \ldots, u_k, \; a_0, \ldots, a_k$ and $b_0, \ldots, b_k$. Also, let $\sum_{k=0}^{\infty} a_k < \infty$ and $\sum_{k=0}^{\infty} b_k < \infty$ a.s. Then, we have $\lim_{k\to\infty} v_k = v$ for a random variable $v \ge 0$ a.s., and $\sum_{k=0}^{\infty} u_k < \infty$ a.s.*

In the following lemma, we show a relation of $p_{i,k}$ and $x_{i,k-1}$ associated with any convex function $h$ which will be

often used in the analysis. For example, $h(x) = \|x - a\|^2$ for some $a \in \mathbb{R}^n$ or $h(x) = \text{dist}^2(x, \mathcal{X})$.

*Lemma 3: Let Assumption 1 hold. Then, for any convex function $h : \mathbb{R}^n \to \mathbb{R}$, we have*

$$\sum_{i \in \mathcal{V}} h(p_{i,k}) \leq \sum_{i \in \mathcal{V}} h(x_{i,k-1})$$

*Proof:* The doubly stochasticity of the weights plays a crucial role in this lemma. From the definition of $p_{i,k}$ in (7a),

$$\begin{aligned}
\sum_{i \in \mathcal{V}} h(p_{i,k}) &\leq \sum_{i \in \mathcal{V}} \sum_{j \in \mathcal{V}} [W_k]_{ij} h(x_{j,k-1}) \\
&= \sum_{j \in \mathcal{V}} \left( \sum_{i \in \mathcal{V}} [W_k]_{ij} \right) h(x_{j,k-1}) \\
&= \sum_{j \in \mathcal{V}} h(x_{j,k-1}).
\end{aligned}$$

■

Lastly, for the convergence proof of our algorithm, we use a result from [52] which shows the averaged iterates can still arrive at consensus if the errors behave nicely.

*Lemma 4: Let Assumptions 1 and 5 hold. Consider the iterates generated by*

$$\theta_{i,k} = \sum_{j \in \mathcal{V}} [W_k]_{ij} \theta_{j,k-1} + e_{i,k}, \text{ for } i \in \mathcal{V}. \quad (14)$$

*Let $\bar{\theta}_k$ denote the average of $\theta_{i,k}$ for $i \in \mathcal{V}$, i.e., $\bar{\theta}_k = \frac{1}{N} \sum_{i \in \mathcal{V}} \theta_{i,k}$. Suppose there exists a nonnegative nonincreasing scalar sequence $\{\alpha_k\}$ such that*

$$\sum_{k=1}^{\infty} \alpha_k \|e_{i,k}\| < \infty, \text{ for all } i \in \mathcal{V}.$$

*Then, for all $i, j \in \mathcal{V}$*

$$\sum_{k=1}^{\infty} \alpha_k \|\theta_{i,k} - \theta_{j,k}\| < \infty.$$

*Furthermore, for all $i \in \mathcal{V}$ and $k \geq 1$,*

$$\|\theta_{i,k} - \bar{\theta}_k\| \leq N \gamma \beta^k \max_j \|\theta_{j,0}\|$$

$$+ \gamma \sum_{\ell=0}^{k-1} \beta^{k-\ell} \sum_{j=1}^{N} \|e_{j,\ell+1}\| + \frac{1}{N} \sum_{j=1}^{N} \|e_{j,k}\| + \|e_{i,k}\|$$

*where $\gamma$ and $\beta$ are defined as*

$$\gamma = \left(1 - \frac{\nu}{2N^2}\right)^{-2}, \quad \beta = \left(1 - \frac{\nu}{2N^2}\right)^{\frac{1}{Q}}.$$

## B. Lemmas

We need a series of lemmas for proving Proposition 1 and 2. We first state an auxiliary lemma that will be later used to relate two consecutive iterates $x_{i,k}$ and $x_{i,k-1}$. This lemma can be shown by combining two existing results in [53] and [27], but we include it here for completeness.

*Lemma 5: Let Assumptions 2 and 3 hold. Let the iterates $\{p_{i,k}\}$, $\{v_{i,k}\}$ and $\{x_{i,k}\}$ be generated by the algorithm (7a)-(7c). Then, we have almost surely for any $\check{x}, z \in \mathcal{X}_0$, $i \in \mathcal{V}$ and $k \geq 1$,*

$$\|x_{i,k} - \check{x}\|^2 \leq \|p_{i,k} - \check{x}\|^2 - 2\alpha_k (f_i(z) - f_i(\check{x}))$$

$$- \frac{\tau - 1}{\tau C_g^2} \left( g^+(p_{i,k}, \omega_{i,k}) \right)^2$$

$$+ \frac{1}{4\eta} \|p_{i,k} - z\|^2 + D_{\tau,\eta} \alpha_k^2,$$

*where $D_{\tau,\eta} = (\tau + 4\eta + 1) C_{f_i}^2$ and $\eta, \tau > 0$ are arbitrary.*

*Proof:* In the light of [53, Theorem 1], we obtain from algorithm (7c) and Assumption 2(e)

$$\|x_{i,k} - \check{x}\|^2 \leq \|v_{i,k} - \check{x}\|^2 - \frac{(g^+(v_{i,k}, \omega_{i,k}))^2}{C_g^2}, \quad (15)$$

for any $\check{x} \in \mathcal{X}_0$. We can rewrite $g^+(v_{i,k}, \omega_{i,k}) = (g^+(v_{i,k}, \omega_{i,k}) - g^+(p_{i,k}, \omega_{i,k})) + g^+(p_{i,k}, \omega_{i,k})$. Therefore,

$$\begin{aligned}
(g^+&(v_{i,k}, \omega_{i,k}))^2 \\
&\geq 2g^+(p_{i,k}, \omega_{i,k}) \left( g^+(v_{i,k}, \omega_{i,k}) - g^+(p_{i,k}, \omega_{i,k}) \right) \\
&+ \left( g^+(p_{i,k}, \omega_{i,k}) \right)^2.
\end{aligned} \quad (16)$$

The first term on the right-hand side of (16) can be further estimated as

$$\begin{aligned}
2g^+&(p_{i,k}, \omega_{i,k}) \left( g^+(v_{i,k}, \omega_{i,k}) - g^+(p_{i,k}, \omega_{i,k}) \right) \\
&\geq -2g^+(p_{i,k}, \omega_{i,k}) \left| g^+(v_{i,k}, \omega_{i,k}) - g^+(p_{i,k}, \omega_{i,k}) \right| \\
&\geq -2C_g \|v_{i,k} - p_{i,k}\| g^+(p_{i,k}, \omega_{i,k}),
\end{aligned} \quad (17)$$

where the last inequality is from relation (13). From the definition in (7b) and Assumption 2(c), we further have that

$$\begin{aligned}
2C_g &\|v_{i,k} - p_{i,k}\| g^+(p_{i,k}, \omega_{i,k}) \\
&\leq 2\alpha_k C_g C_{f_i} g^+(p_{i,k}, \omega_{i,k}) \\
&\leq \tau \alpha_k^2 C_g^2 C_{f_i}^2 + \frac{1}{\tau} \left( g^+(p_{i,k}, \omega_{i,k}) \right)^2,
\end{aligned} \quad (18)$$

where the last inequality is obtained by using $2|a||b| \leq \tau a^2 + \frac{1}{\tau} b^2$ and $\tau > 0$ is arbitrary. Using relations (17)-(18) in (16), we obtain,

$$\begin{aligned}
(g^+&(v_{i,k}, \omega_{i,k}))^2 \\
&\geq -\tau \alpha_k^2 C_g^2 C_{f_i}^2 + \left(1 - \frac{1}{\tau}\right) \left( g^+(p_{i,k}, \omega_{i,k}) \right)^2.
\end{aligned}$$

Hence, for all $\check{x} \in \mathcal{X}_0$,

$$\begin{aligned}
\|x_{i,k} - \check{x}\|^2 \leq &\|v_{i,k} - \check{x}\|^2 - \frac{\tau - 1}{\tau C_g^2} \left( g^+(p_{i,k}, \omega_{i,k}) \right)^2 \\
&+ \tau C_{f_i}^2 \alpha_k^2.
\end{aligned} \quad (19)$$

As the update rule in (7b) coincides with the algorithm in [27], we can reuse another existing lemma [27, Lemma 3]. That is, for any $\check{x}, z \in \mathcal{X}_0$, we have

$$\begin{aligned}
\|v_{i,k} - \check{x}\|^2 \leq &\|p_{i,k} - \check{x}\|^2 - 2\alpha_k (f_i(z) - f_i(\check{x})) \\
&+ \frac{1}{4\eta} \|p_{i,k} - z\|^2 + \alpha_k^2 (1 + 4\eta) C_{f_i}^2,
\end{aligned} \quad (20)$$

where $\eta > 0$ is arbitrary. Substituting this inequality in relation (19) concludes the proof. ■

Since we use an approximate projection, we cannot guarantee the feasibility of the iterates $\{x_{i,k}\}$ and $\{p_{i,k}\}$. In the next lemma, we prove that $\{p_{i,k}\}$ and $\{x_{i,k}\}$ for all $i \in \mathcal{V}$ asymptotically achieve feasibility. To this end, we define the

following quantity: For all $i \in \mathcal{V}$ and $k \geq 1$, $z_{i,k}$ is defined as the projection of $p_{i,k}$ on the feasible set $\mathcal{X}$, i.e.,

$$z_{i,k} \triangleq \Pi_{\mathcal{X}}[p_{i,k}]. \tag{21}$$

*Lemma 6:* Let Assumptions 1 - 4 hold. Let the sequence $\{\alpha_k\}$ be such that $\sum_{k=1}^{\infty} \alpha_k^2 < \infty$. Then, the iterates $\{p_{i,k}\}$ and $\{x_{i,k}\}$ generated by each agent $i \in \mathcal{V}$ via method (7a)-(7c) satisfy:

(a) $\sum_{k=1}^{\infty} \mathrm{dist}^2(p_{i,k}, \mathcal{X}) < \infty \quad a.s.$

(b) $\sum_{k=1}^{\infty} \|x_{i,k} - z_{i,k}\|^2 < \infty \quad a.s.$

where $z_{i,k}$ is defined in (21).

*Proof:* We use Lemma 5 with $\check{x} = z = z_{i,k}$. Therefore, for any $\omega_{i,k} \in \Omega_i$, $i \in \mathcal{V}$, and $k \geq 1$, we obtain

$$\begin{aligned}
&\|x_{i,k} - z_{i,k}\|^2 \\
&\leq \|p_{i,k} - z_{i,k}\|^2 - \frac{\tau - 1}{\tau C_g^2} \left( g^+(p_{i,k}, \omega_{i,k}) \right)^2 \\
&\quad + \frac{1}{4\eta} \|p_{i,k} - z_{i,k}\|^2 + D_{\tau,\eta} \alpha_k^2,
\end{aligned} \tag{22}$$

where $D_{\tau,\eta} = (\tau + 4\eta + 1)C_{f_i}^2$ and $\eta, \tau > 0$ are arbitrary. By the definition of the projection, we have

$$\mathrm{dist}(p_{i,k}, \mathcal{X}) = \|p_{i,k} - z_{i,k}\|, \text{ and}$$
$$\mathrm{dist}(x_{i,k}, \mathcal{X}) = \|x_{i,k} - \Pi_{\mathcal{X}}[x_{i,k}]\| \leq \|x_{i,k} - z_{i,k}\|.$$

Upon substituting these estimates in relation (22), we obtain

$$\begin{aligned}
\mathrm{dist}^2(x_{i,k}, \mathcal{X}) &\leq \mathrm{dist}^2(p_{i,k}, \mathcal{X}) - \frac{\tau - 1}{\tau C_g^2} \left( g^+(p_{i,k}, \omega_{i,k}) \right)^2 \\
&\quad + \frac{1}{4\eta} \mathrm{dist}^2(p_{i,k}, \mathcal{X}) + D_{\tau,\eta} \alpha_k^2.
\end{aligned} \tag{23}$$

Taking the expectation conditioned on $\mathcal{F}_{k-1}$ and noting that $p_{i,k}$ is fully determined by $\mathcal{F}_{k-1}$, we have almost surely for any $i \in \mathcal{V}$ and $k \geq 1$

$$\begin{aligned}
&\mathsf{E}\left[\mathrm{dist}^2(x_{i,k}, \mathcal{X}) \mid \mathcal{F}_{k-1}\right] \\
&\leq \mathrm{dist}^2(p_{i,k}, \mathcal{X}) - \frac{\tau - 1}{\tau C_g^2} \mathsf{E}\left[ \left( g^+(p_{i,k}, \omega_{i,k}) \right)^2 \mid \mathcal{F}_{k-1} \right] \\
&\quad + \frac{1}{4\eta} \mathrm{dist}^2(p_{i,k}, \mathcal{X}) + D_{\tau,\eta} \alpha_k^2.
\end{aligned} \tag{24}$$

Furthermore, choosing $\tau = 4$, $\eta = cC_g^2$ and using Assumption 4 yield

$$\begin{aligned}
\mathsf{E}\left[ \mathrm{dist}^2(x_{i,k}, \mathcal{X}) \mid \mathcal{F}_{k-1} \right] &\leq \mathrm{dist}^2(p_{i,k}, \mathcal{X}) \\
&\quad - \frac{1}{2cC_g^2} \mathrm{dist}^2(p_{i,k}, \mathcal{X}) + (5 + 4cC_g^2)C_{f_i}^2 \alpha_k^2.
\end{aligned} \tag{25}$$

Finally, by summing over all $i$ and using Lemma 3 with $h(x) = \mathrm{dist}^2(x, \mathcal{X})$, we arrive at the following relation:

$$\begin{aligned}
\sum_{i \in \mathcal{V}} \mathsf{E}\left[ \mathrm{dist}^2(x_{i,k}, \mathcal{X}) \mid \mathcal{F}_{k-1} \right] &\leq \sum_{i \in \mathcal{V}} \mathrm{dist}^2(x_{i,k-1}, \mathcal{X}) \\
&\quad - \frac{1}{2cC_g^2} \sum_{i \in \mathcal{V}} \mathrm{dist}^2(p_{i,k}, \mathcal{X}) + DN\alpha_k^2,
\end{aligned} \tag{26}$$

where $D = (5 + 4cC_g^2)C_f^2$ and $C_f = \max_{i \in \mathcal{V}} C_{f_i}$. Therefore, for all $k \geq 1$, all the conditions of the convergence theorem (Theorem 1) are satisfied and we conclude that

$$\sum_{k=1}^{\infty} \mathrm{dist}^2(p_{i,k}, \mathcal{X}) < \infty \quad \text{for all } i \in \mathcal{V} \quad a.s. \tag{27}$$

Lastly, from relation (22) and the chosen values for $\tau$ and $\eta$, we obtain for any $i \in \mathcal{V}$ and $k \geq 1$ almost surely

$$\|x_{i,k} - z_{i,k}\|^2 \leq \left( 1 + \frac{1}{4cC_g} \right) \mathrm{dist}^2(p_{i,k}, \mathcal{X}) + D\alpha_k^2.$$

Therefore, in view of the result in (27) and $\sum_{k=1}^{\infty} \alpha_k^2 < \infty$, the relation above implies

$$\sum_{k=1}^{\infty} \|x_{i,k} - z_{i,k}\|^2 < \infty \quad \text{for all } i \in \mathcal{V} \quad a.s.,$$

which is our desired result. ∎

To complete the proof, we show in part (a) of the next lemma that the error $e_{i,k}$ due to the perturbations made after the consensus step (7a), i.e.,

$$e_{i,k} \triangleq x_{i,k} - p_{i,k}, \tag{28}$$

eventually converges to zero for all $i \in \mathcal{V}$. This will allow us to invoke Lemma 4 and show the iterate consensus. In part (b) of the next lemma, we show that the sequences $\{z_{i,k}\}$ arrive at consensus by converging to their mean $\bar{z}_k$, i.e., for $k \geq 1$

$$\bar{z}_k \triangleq \frac{1}{N} \sum_{i \in \mathcal{V}} z_{i,k}. \tag{29}$$

In part (c) of the next lemma, we show the network error term is summable.

*Lemma 7:* Let Assumptions 1-5 hold. Let the sequence $\{\alpha_k\}$ be nonnegative nonincreasing and $\sum_{k=1}^{\infty} \alpha_k^2 < \infty$. Then, we have for all $i \in \mathcal{V}$

(a) $\sum_{k=1}^{\infty} \|e_{i,k}\|^2 < \infty \quad a.s.$

(b) $\lim_{k \to \infty} \|z_{i,k} - \bar{z}_k\| = 0 \quad a.s.$

(c) $\sum_{k=1}^{\infty} \alpha_k \|z_{i,k} - \bar{z}_k\| < \infty \quad a.s.$

*Proof: Part (a):* From the relation (7a)-(7c), $e_{i,k}$ in (28) can be viewed as the perturbation that we make on $p_{i,k}$ after the network consensus step (7a). Consider $\|e_{i,k}\|$, for which we can write

$$\|e_{i,k}\| \leq \|x_{i,k} - z_{i,k}\| + \|z_{i,k} - p_{i,k}\|.$$

Applying $(a + b)^2 \leq 2a^2 + 2b^2$ in the above inequality, we have

$$\|e_{i,k}\|^2 \leq 2\|x_{i,k} - z_{i,k}\|^2 + 2\mathrm{dist}^2(p_{i,k}, \mathcal{X}).$$

Summing this over $k$ and using Lemma 6, we obtain the desired result.

*Part (b):* By applying the inequality $2ab \leq a^2 + b^2$ to each term in $\alpha_k \|e_{i,k}\|$ and using Lemma 7(a), we further obtain for all $i \in \mathcal{V}$

$$\sum_{k=1}^{\infty} \alpha_k \|e_{i,k}\| \leq \frac{1}{2} \sum_{k=1}^{\infty} \alpha_k^2 + \frac{1}{2} \sum_{k=1}^{\infty} \|e_{i,k}\|^2 < \infty \quad a.s. \tag{30}$$

Using the relation above, (7a) and $x_{i,k} = p_{i,k} + e_{i,k}$, we can invoke Lemma 4 with $\theta_{i,k} = x_{i,k}$. Therefore, it follows that

$$\sum_{k=1}^{\infty} \alpha_k \|x_{i,k} - x_{j,k}\| < \infty \text{ for all } i, j \in \mathcal{V} \quad a.s. \quad (31)$$

Furthermore, for all $i \in \mathcal{V}$ and $k \geq 1$,

$$\|x_{i,k} - \bar{x}_k\| \leq N\gamma\beta^k \max_j \|x_{j,0}\| \quad (32)$$

$$+ \gamma \sum_{\ell=0}^{k-1} \beta^{k-\ell} \sum_{j=1}^{N} \|e_{j,\ell+1}\| + \frac{1}{N} \sum_{j=1}^{N} \|e_{j,k}\| + \|e_{i,k}\|.$$

From the fact that $0 < \beta < 1$ and part(a), we know the first term and the last two terms on the right-hand side converge to zero. To show the second term also converges to zero, we will use the following result from [54, Lemma 3.1(a)].

*Lemma 8:* Let $\zeta_k$ be a scalar sequence. If $\lim_{k\to\infty} \zeta_k = \zeta$ and $0 < \beta < 1$, then $\lim_{k\to\infty} \sum_{\ell=0}^{k-1} \beta^{k-\ell} \zeta_\ell = \frac{\zeta}{1-\beta}$.

From this lemma and the result in part(a), we know that the second term on the right-hand side of (32) also converges to zero. Therefore, we have for all $i \in \mathcal{V}$

$$\lim_{k\to\infty} \|x_{i,k} - \bar{x}_k\| = 0. \quad (33)$$

We next consider the term $\|z_{i,k} - \bar{z}_k\|$, for which by using $\bar{z}_k = \frac{1}{N} \sum_{\ell \in \mathcal{V}} z_{\ell,k}$ we have

$$\|z_{i,k} - \bar{z}_k\| = \left\| \frac{1}{N} \sum_{\ell \in \mathcal{V}} (z_{i,k} - z_{\ell,k}) \right\|$$

$$\leq \frac{1}{N} \sum_{\ell \in \mathcal{V}} \|z_{i,k} - z_{\ell,k}\| \leq \frac{1}{N} \sum_{\ell \in \mathcal{V}} \|p_{i,k} - p_{\ell,k}\|,$$

where the first inequality is obtained by the convexity of the norm and the last inequality follows by the non-expansive projection property in Lemma 2. Furthermore, by using $\|p_{i,k} - p_{\ell,k}\| \leq \|p_{i,k} - \bar{p}_k\| + \|p_{\ell,k} - \bar{p}_k\|$, we obtain for every $i \in \mathcal{V}$

$$\|z_{i,k} - \bar{z}_k\| \leq \|p_{i,k} - \bar{p}_k\| + \frac{1}{N} \sum_{\ell \in \mathcal{V}} \|p_{\ell,k} - \bar{p}_k\|. \quad (34)$$

We next consider $\|p_{i,k} - \bar{p}_k\|$. By using the doubly stochasticity of $W_k$, convexity of the norm and the fact that $0 \leq [W_k]_{ij} \leq 1$, we obtain

$$\|p_{i,k} - \bar{p}_k\| \leq \sum_{j \in \mathcal{V}} [W_k]_{ij} \|x_{j,k-1} - \bar{p}_k\|$$

$$\leq \sum_{j \in \mathcal{V}} \left\| x_{j,k-1} - \frac{1}{N} \sum_{\ell \in \mathcal{V}} x_{\ell,k-1} \right\|,$$

where in the last equality we use $\bar{p}_k = \frac{1}{N} \sum_{\ell \in \mathcal{V}} \sum_{j \in \mathcal{V}} [W_k]_{\ell j} x_{j,k-1} = \frac{1}{N} \sum_{\ell \in \mathcal{V}} x_{\ell,k-1}$. Therefore, by using the convexity of the norm again, we see

$$\|p_{i,k} - \bar{p}_k\| \leq \frac{1}{N} \sum_{j \in \mathcal{V}} \sum_{\ell \in \mathcal{V}} \|x_{j,k-1} - x_{\ell,k-1}\| \quad (35)$$

$$\leq \frac{1}{N} \sum_{j \in \mathcal{V}} \sum_{\ell \in \mathcal{V}} \left( \|x_{j,k-1} - \bar{x}_{k-1}\| + \|x_{\ell,k-1} - \bar{x}_{k-1}\| \right).$$

Combining this relation with (34) and using the result in (33), we obtain the desired result.

*Part (c):* By using relation (31) in (35), we obtain

$$\sum_{k=1}^{\infty} \alpha_k \|p_{i,k} - \bar{p}_k\|$$

$$\leq \sum_{k=1}^{\infty} \frac{\alpha_k}{N} \sum_{j \in \mathcal{V}} \sum_{\ell \in \mathcal{V}} \|x_{j,k-1} - x_{\ell,k-1}\| < \infty. \quad (36)$$

Upon summing the relation (34) over $i \in \mathcal{V}$, we find

$$\sum_{i \in \mathcal{V}} \|z_{i,k} - \bar{z}_k\| \leq 2 \sum_{i \in \mathcal{V}} \|p_{i,k} - \bar{p}_k\|. \quad (37)$$

Therefore, from (36) and (37), we obtain

$$\sum_{i \in \mathcal{V}} \sum_{k=1}^{\infty} \alpha_k \|z_{i,k} - \bar{z}_k\| \leq 2 \sum_{i \in \mathcal{V}} \sum_{k=1}^{\infty} \alpha_k \|p_{i,k} - \bar{p}_k\| < \infty,$$

which is the desired result. ∎

In the next lemma, we use standard convexity analysis to lower-bound the term $\sum_{i \in \mathcal{V}}(f_i(z_{i,k}) - f_i(\check{x}))$ with a network error term and a global term.

*Lemma 9: Let Assumption 2 hold. Then, for all $\check{x} \in \mathcal{X}$, we have*

$$\sum_{i \in \mathcal{V}} (f_i(z_{i,k}) - f_i(\check{x})) \geq -C_f \sum_{i \in \mathcal{V}} \|z_{i,k} - \bar{z}_k\| + f(\bar{z}_k) - f(\check{x}),$$

*where $C_f = \max_{i \in \mathcal{V}} C_{f_i}$.*

*Proof:* Recall that $f(x) = \sum_{i=1}^{m} f_i(x)$. Recall that $\bar{z}_k = \frac{1}{N} \sum_{\ell \in \mathcal{V}} z_{\ell,k}$. Using $\bar{z}_k$ and $f$, we can rewrite the term $f_i(z_{i,k}) - f_i(\bar{x})$ as follows:

$$\sum_{i \in \mathcal{V}} (f_i(z_{i,k}) - f_i(\check{x}))$$

$$= \sum_{i \in \mathcal{V}} (f_i(z_{i,k}) - f_i(\bar{z}_k)) + (f(\bar{z}_k) - f(\check{x})). \quad (38)$$

Furthermore, using the convexity of each function $f_i$, we obtain

$$\sum_{i \in \mathcal{V}} (f_i(z_{i,k}) - f_i(\bar{z}_k)) \geq \sum_{i \in \mathcal{V}} \langle s_{i,k}, z_{i,k} - \bar{z}_k \rangle$$

$$\geq -\sum_{i \in \mathcal{V}} \|s_{i,k}\| \|z_{i,k} - \bar{z}_k\|,$$

where $s_{i,k}$ is a subgradient of $f_i$ at $\bar{z}_k$. Since $\bar{z}_k$ is a convex combination of points $z_{i,k} \in \mathcal{X} \subseteq \mathcal{X}_0$, it follows that $\bar{z}_k \in \mathcal{X}_0$. This observation and Assumption 2(c), stating that the subgradients of $f_i(x)$ are uniformly bounded for $x \in \mathcal{X}_0$, yield

$$\sum_{i \in \mathcal{V}} (f_i(z_{i,k}) - f_i(\bar{z}_k)) \geq -C_f \sum_{i \in \mathcal{V}} \|z_{i,k} - \bar{z}_k\|, \quad (39)$$

where $C_f = \max_{i \in \mathcal{V}} C_{f_i}$. Therefore, from (38) and (39), we have that

$$\sum_{i \in \mathcal{V}} (f_i(z_{i,k}) - f_i(\check{x}))$$

$$\geq -C_f \sum_{i \in \mathcal{V}} \|z_{i,k} - \bar{z}_k\| + f(\bar{z}_k) - f(\check{x}).$$

∎

## C. Proof of Proposition 1

We invoke Lemma 5 with $z = z_{i,k} = \Pi_{\mathcal{X}}[p_{i,k}]$, $\tau = 4$ and $\eta = cC_g^2$. We also let $\check{x} = x^*$ for an arbitrary $x^* \in \mathcal{X}^*$. Therefore, for any $x^* \in \mathcal{X}^*$, $i \in \mathcal{V}$ and $k \geq 1$, we almost surely have

$$
\begin{aligned}
\|x_{i,k}-x^*\|^2 \leq & \|p_{i,k} - x^*\|^2 - 2\alpha_k(f_i(z_{i,k}) - f_i(x^*)) \quad (40)\\
& - \frac{3}{4C_g^2} \left(g^+(p_{i,k}, \omega_{i,k})\right)^2\\
& + \frac{1}{4cC_g^2}\text{dist}^2(p_{i,k}, \mathcal{X}) + (5 + 4cC_g^2)C_{f_i}^2\alpha_k^2.
\end{aligned}
$$

By taking the expectation conditioned on $\mathcal{F}_{k-1}$ in the above relation and summing this over $i \in \mathcal{V}$, we obtain

$$
\begin{aligned}
& \sum_{i\in\mathcal{V}}\mathsf{E}\left[\|x_{i,k} - x^*\|^2 \mid \mathcal{F}_{k-1}\right]\\
& \leq \sum_{i\in\mathcal{V}}\|p_{i,k} - x^*\|^2 - 2\alpha_k\sum_{i\in\mathcal{V}}(f_i(z_{i,k}) - f_i(x^*))\\
& \quad - \frac{3}{4C_g^2}\sum_{i\in\mathcal{V}}\mathsf{E}\left[\left(g^+(p_{i,k}, \omega_{i,k})\right)^2 \mid \mathcal{F}_{k-1}\right]\\
& \quad + \frac{1}{4cC_g^2}\sum_{i\in\mathcal{V}}\text{dist}^2(p_{i,k}, \mathcal{X}) + DN\alpha_k^2,
\end{aligned}
$$

where $D = (5 + 4cC_g^2)C_f^2$ with $C_f = \max_{i\in\mathcal{V}} C_{f_i}$. Now we use Lemma 3 with $h(x) = \|x - x^*\|^2$, Assumption 4 and Lemma 9 with $\check{x} = x^*$ to further estimate the terms on the right-hand side. From these, obtain almost surely for any $k \geq 1$ and $x^* \in \mathcal{X}^*$,

$$
\begin{aligned}
& \sum_{i\in\mathcal{V}}\mathsf{E}\left[\|x_{i,k} - x^*\|^2 \mid \mathcal{F}_{k-1}\right]\\
& \leq \sum_{i\in\mathcal{V}}\|x_{i,k-1} - x^*\|^2 - 2\alpha_k\sum_{i\in\mathcal{V}}(f(\bar{z}_k) - f(x^*))\\
& \quad - \frac{1}{2cC_g^2}\sum_{i\in\mathcal{V}}\text{dist}^2(p_{i,k}, \mathcal{X})\\
& \quad + 2\alpha_k C_f\sum_{i\in\mathcal{V}}\|z_{i,k} - \bar{z}_k\| + DN\alpha_k^2.
\end{aligned}
$$

Since $\bar{z}_k \in \mathcal{X}$, we have $f(\bar{z}_k) - f(x^*) \geq 0$. Thus, under the assumption $\sum_{k=0}^\infty \alpha_k^2 < \infty$ and Lemma 7(c), the above relation satisfies all the conditions of the convergence Theorem 1. Using this theorem, we have the following results.

*Result 1:* The sequence $\{\sum_{i\in\mathcal{V}}\|x_{i,k} - x^*\|\}$ is convergent *a.s.* for every $x^* \in \mathcal{X}^*$.

*Result 2:* For every $x^* \in \mathcal{X}^*$,

$$
\sum_{k=1}^\infty \alpha_k(f(\bar{z}_k) - f(x^*)) < \infty \quad a.s.
$$

From *Result 1* and Lemma 6(b), we know that the sequence $\{\sum_{i\in\mathcal{V}}\|z_{i,k} - x^*\|\}$ is convergent *a.s.* for every $x^* \in \mathcal{X}^*$. This and Lemma 7(b) imply that $\|\bar{z}_k - x^*\|$ is also convergent *a.s.* for every $x^* \in \mathcal{X}^*$. From *Result 2*, $\sum_{k=1}^\infty \alpha_k = \infty$, and the continuity of $f$, it follows that the sequence $\{\bar{z}_k\}$ must have one accumulation point in the set $\mathcal{X}^*$ *a.s.* This and the fact that $\{\|\bar{z}_k - x^*\|\}$ is convergent *a.s.* for every $x^* \in \mathcal{X}^*$ imply that for a random point $x^\star \in \mathcal{X}^*$,

$$
\lim_{k\to\infty}\bar{z}_k = x^\star \quad a.s. \quad (41)
$$

We now prove the following claim: For all $i \in \mathcal{V}$

$$
\lim_{k\to\infty}x_{i,k} = x^\star \quad a.s. \quad (42)
$$

Consider

$$
\|x_{i,k} - x^\star\| \leq \|x_{i,k} - z_{i,k}\| + \|z_{i,k} - \bar{z}_k\| + \|\bar{z}_k - x^\star\|.
$$

From Lemma 6(b), Lemma 7(b) and (41), all the terms on the right-hand side converge to zero *a.s.* Therefore, it is obvious that claim (42) holds, which is our desired result.

## D. Proof of Proposition 2

The line of proof is similar to that in Proposition 1. Therefore, we only lay down the differences.

Using the definition of $\pi$ in Lemma 1, we have

$$
[\pi]_j = \sum_{i\in\mathcal{V}}[\pi]_i[W]_{ij}, \quad \text{for all } j \in \mathcal{V}. \quad (43)
$$

Also, in the proof we consider the following weighted averages rather than the true averages $\bar{x}_k$, $\bar{p}_k$ and $\bar{z}_k$.

$$
\hat{x}_k \triangleq \sum_{i\in\mathcal{V}}[\pi]_i x_{i,k}, \quad \hat{p}_k \triangleq \sum_{i\in\mathcal{V}}[\pi]_i p_{i,k}, \quad (44)
$$
$$
\text{and } \hat{z}_k \triangleq \sum_{i\in\mathcal{V}}[\pi]_i z_{i,k}.
$$

First, notice that Lemma 5 still holds in this case as it does not require Assumption 1.

*Changes in Lemma 4:* Combining with the results in [5], [16], [55]–[58], Lemma 4 still holds in this case by replacing $\bar{\theta}_k$ with $\hat{\theta}_k \triangleq \sum_{i\in\mathcal{V}}[\pi]_i\theta_{i,k}$ and re-defining the constants $\gamma$ and $\beta$ as

$$
\gamma = 2, \quad \beta = 1 - \frac{1}{N^N}.
$$

If in addition the underlying graph $\mathcal{G}$ is regular, then we have

$$
\gamma = \sqrt{2}, \quad \beta = \min\left\{1 - \frac{1}{4N^3}, \ \max_{k\geq 1}\sigma_2(W)\right\},
$$

where $\sigma_2(W)$ is the second largest singular value of $W$.

*Changes in Lemma 6:* By multiplying $[\pi]_i$ to (25) and summing over $i \in \mathcal{V}$, we obtain

$$
\begin{aligned}
\sum_{i\in\mathcal{V}}\mathsf{E}\left[[\pi]_i\text{dist}^2(x_{i,k}, \mathcal{X}) \mid \mathcal{F}_{k-1}\right] \leq & \sum_{i\in\mathcal{V}}[\pi]_i\text{dist}^2(p_{i,k}, \mathcal{X})\\
& - \frac{1}{2cC_g^2}\sum_{i\in\mathcal{V}}[\pi]_i\text{dist}^2(p_{i,k}, \mathcal{X}) + DN\alpha_k^2. \quad (45)
\end{aligned}
$$

From the definition of $p_{i,k}$ in (7a) and the convexity of the distance function, we have

$$
\begin{aligned}
\sum_{i\in\mathcal{V}}[\pi]_i\text{dist}^2(p_{i,k}, \mathcal{X}) & \leq \sum_{i\in\mathcal{V}}\sum_{j\in\mathcal{V}}[\pi]_i[W]_{ij}\text{dist}^2(x_{j,k-1}, \mathcal{X})\\
& \leq \sum_{j\in\mathcal{V}}[\pi]_j\text{dist}^2(x_{j,k-1}, \mathcal{X}),
\end{aligned}
$$

where the last inequality follows from (43). Combining this result with (45), we obtain

$$\sum_{i \in \mathcal{V}} \mathsf{E}\left[[\pi]_i \mathrm{dist}^2(x_{i,k}, \mathcal{X}) \mid \mathcal{F}_{k-1}\right] \leq \sum_{i \in \mathcal{V}} [\pi]_i \mathrm{dist}^2(x_{i,k-1}, \mathcal{X})$$
$$- \frac{1}{2cC_g^2} \sum_{i \in \mathcal{V}} [\pi]_i \mathrm{dist}^2(p_{i,k}, \mathcal{X}) + DN\alpha_k^2,$$

in which all the conditions of Theorem 1 holds. Hence, all the remaining results follow immediately.

*Changes in Lemma 7:* All the results still hold by replacing $\bar{z}_k$, $\bar{p}_k$ and $\bar{x}_k$ with $\hat{z}_k$, $\hat{p}_k$ and $\hat{x}_k$, respectively. Especially, from relation (43) we have

$$\hat{p}_k = \sum_{i \in \mathcal{V}} [\pi]_i \sum_{j \in \mathcal{V}} [W]_{ij} x_{j,k-1} = \sum_{j \in \mathcal{V}} [\pi]_j x_{j,k-1} = \hat{x}_{k-1},$$

and all the results follow immediately.

*Changes in Proposition 1:* By multiplying $[\pi]_i$ to (40), summing this over $i \in \mathcal{V}$ and considering $f_i(x) = \frac{1}{N} a^\top x$, we have

$$\sum_{i \in \mathcal{V}} [\pi]_i \|x_{i,k} - x^*\|^2 \leq \sum_{i \in \mathcal{V}} [\pi]_i \|p_{i,k} - x^*\|^2$$
$$- \frac{2\alpha_k}{N} a^\top (\hat{z}_k - x^*) - \frac{3}{4C_g^2} \sum_{i \in \mathcal{V}} [\pi]_i \left(g^+(p_{i,k}, \omega_{i,k})\right)^2$$
$$+ \frac{1}{4cC_g^2} \sum_{i \in \mathcal{V}} [\pi]_i \mathrm{dist}^2(p_{i,k}, \mathcal{X}) + DN\alpha_k^2,$$

where we used the fact that $\sum_{i \in \mathcal{V}} [\pi]_i z_{i,k} = \hat{z}_k$.

Now we use (43) and Assumption 4 to obtain almost surely for any $k \geq 1$ and $x^* \in \mathcal{X}^*$,

$$\sum_{i \in \mathcal{V}} \mathsf{E}\left[[\pi]_i \|x_{i,k} - x^*\|^2 \mid \mathcal{F}_{k-1}\right]$$
$$\leq \sum_{i \in \mathcal{V}} [\pi]_i \|x_{i,k-1} - x^*\|^2 - \frac{2\alpha_k}{N} a^\top (\hat{z}_k - x^*)$$
$$- \frac{1}{2cC_g^2} \sum_{i \in \mathcal{V}} [\pi]_i \mathrm{dist}^2(p_{i,k}, \mathcal{X}) + DN\alpha_k^2.$$

Since $\hat{z}_k \in \mathcal{X}$, we have $a^\top (\hat{z}_k - x^*) \geq 0$. Thus, under the assumption $\sum_{k=0}^{\infty} \alpha_k^2 < \infty$ and Lemma 6(a), the above relation satisfies all the conditions of the convergence Theorem 1.

Using this theorem, we have the following results.

*Result 1:* The sequence $\{\sum_{i \in \mathcal{V}} [\pi]_i \|x_{i,k} - x^*\|\}$ is convergent *a.s.* for every $x^* \in \mathcal{X}^*$.

*Result 2:* For every $x^* \in \mathcal{X}^*$,

$$\sum_{k=1}^{\infty} \alpha_k a^\top (\hat{z}_k - x^*) < \infty \quad a.s.$$

From *Result 1* and Lemma 6(b), we know that the sequence $\{\sum_{i \in \mathcal{V}} [\pi]_i \|z_{i,k} - x^*\|\}$ is convergent *a.s.* for every $x^* \in \mathcal{X}^*$. This and Lemma 7(b) imply that $\|\hat{z}_k - x^*\|$ is also convergent *a.s.* for every $x^* \in \mathcal{X}^*$. From *Result 2*, $\sum_{k=1}^{\infty} \alpha_k = \infty$, it follows that the sequence $\{\hat{z}_k\}$ must have one accumulation point in the set $\mathcal{X}^*$ *a.s.* This and the fact that $\{\|\hat{z}_k - x^*\|\}$ is

convergent *a.s.* for every $x^* \in \mathcal{X}^*$ imply that for a random point $x^\star \in \mathcal{X}^*$,

$$\lim_{k \to \infty} \hat{z}_k = x^\star \quad a.s.$$

The remaining results follow immediately.

## V. SIMULATION RESULTS

In this section, we provide a numerical example showing the effectiveness of the proposed decentralized approximate projection algorithm. We consider optimal gossip averaging which is an example of decentralized optimization.

In many decentralized algorithms, gossip based communication protocols are often used. In these communication protocols, only one agent randomly wakes up at a time (say agent $i$) and selects one of its neighbors (say agent $j$) with probability $p_{ij}$. Then, the two agents exchange their current information through the link $(i, j)$ and take the average. Let $A(i, j)$ denote the averaging matrix associated with the link $(i, j)$. For example, the averaging matrix $A(1, 2)$ of a 4-agent network system looks like

$$A(1, 2) = \begin{bmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Note that the expectation of the averaging matrix $A$ can be represented as $\mathsf{E}[A] = \frac{1}{N} \sum_{i,j \in \mathcal{V}} p_{ij} A(i, j)$.

Let $P$ denote the probability matrix whose component of the $i$-th row and $j$-th column is $p_{ij}$. Our goal here is to find an optimal probability matrix $P^*$ associated with the current communication graph, which is time-invariant and connected, in a decentralized fashion. The convergence speed of the gossip protocol is known to be inversely proportional to $\lambda_2(\mathsf{E}(A))$, which is the second largest eigenvalue of the expected averaging matrix $\mathsf{E}[A]$ (see [59]). Thus, the optimization problem of finding the fastest averaging distribution $P^*$ can be formulated as the following SDP:

$$\min_{s, P} \quad s \tag{46a}$$
$$\text{s.t.} \quad \sum_{i,j \in \mathcal{V}} p_{ij} A(i, j) - \mathbf{1}\mathbf{1}^\top \preceq sI \tag{46b}$$
$$p_{ij} \geq 0, \quad p_{ij} = 0 \text{ if } (i, j) \notin \mathcal{E} \tag{46c}$$
$$\sum_{j \in \mathcal{V}} p_{ij} = 1, \text{ for } i \in \mathcal{V}. \tag{46d}$$

An optimal $P^*$ of the problem (46a)-(46d) computed in a centralized fashion is not useful as gossip protocol is usually required in a decentralized setting. A decentralized method has been proposed in [59], but the method only finds a suboptimal solution. Using our proposed algorithm, we can find the optimal solution of (46a)-(46d) in a decentralized way.

With a slight abuse of notation, let $x \triangleq [s; P]$. In this problem, all agents share the same local objective function, i.e., $f_i(x) = s$ for all $i \in \mathcal{V}$, whereas each agent $i$ has a local constraint set $\mathcal{X}_i = \mathcal{X}_i^1 \bigcap \mathcal{X}_i^2$ where

$$\mathcal{X}_i^1 = \{x \mid \sum_{i,j \in \mathcal{V}} p_{ij} A(i, j) - \mathbf{1}\mathbf{1}^\top \preceq sI\},$$

TABLE I
NUMBER OF ITERATIONS FOR ALL NETWORK AGENTS TO CONVERGE
WITHIN 0.01% OF THE GLOBAL AVERAGE

|          | clique | cycle | star   |
|----------|--------|-------|--------|
| $N = 4$  | 2,170  | 2,819 | 7,190  |
| $N = 15$ | 2,179  | 8,280 | 18,541 |

$$\mathcal{X}_i^2 = \{x \mid \sum_{j \in \mathcal{V}} p_{ij} = 1, \ p_{ij} \geq 0, \ p_{ij} = 0 \text{ if } j \notin \mathcal{N}_i\}.$$

At each iteration of our algorithm, we randomly select a component constraint from $\mathcal{X}_i$ and make a projection. More specifically, we approximate the projection onto the SDP constraint $\mathcal{X}_i^1$ using the equation (11). Note that the constraints (46c)-(46d), which are distributed among agents, will guarantee the structure of the underlying communication graph. Therefore, agents do not require knowledge on the whole graph structure.

We note that due to the compactness of the set $\mathcal{X}_i^2$, the problem (46a)-(46d) satisfies Assumption 2 and the optimal solution set $\mathcal{X}^*$ is nonempty. Assumptions 1, 3 and 5 can be satisfied by construction. Assumption 4 is also satisfied as all inequalities are affine in this case.

We let all agents terminate if their solution is within 0.01% of the global average and the total feasibility violation is less than 0.001. We say the algorithm has converged only when all network agents terminate. Note that this global average based criterion is just used for the sake of simulations. Also, due to the randomness of our algorithm, we repeat all the simulations for 10 times and report their averages.

Table I summarizes the simulation results. It shows the number of iterations until convergence for different numbers of agents ($N$) and underlying communication topologies ($\mathcal{G}_k$). In the experiment, we use $\alpha_{i,k} = \frac{1}{k}$ for all $i \in \mathcal{V}$, and $4, 15$ agents with three different network topologies, namely clique, cycle and star. Note that for this problem the underlying network must be time-invariant, i.e., $\mathcal{G}_k = \mathcal{G}$ for $k \geq 1$, as the gossip algorithm in [59] is built on a fixed undirected graph. As expected, the star graph takes the most iterations for both $N = 4, 15$. Also, when there are more agents in the network, the algorithm takes more iterations.

## VI. CONCLUSION

We have studied a distributed optimization problem defined on a multiagent network which involves nontrivial constraints like LMIs. We have proposed a decentralized algorithm based on random feasibility updates, where we approximate the projection with an additional subgradient step. The proposed algorithm is efficiently applicable for solving any distributed optimization problems which involve lots of computationally prohibitive constraints, for example, decentralized SDPs. We have established the almost sure convergence of our method under two different assumptions on the weight matrices, namely doubly stochastic $\{W_k\}$ over a $Q$-strongly connected sequence of digraphs and row stochastic $W$ over a strongly connected digraph. We have performed experiments on an

optimal gossip averaging problem to verify the performance and convergence of the proposed algorithm.

## REFERENCES

[1] N. Cesa-Bianchi and G. Lugosi, *Prediction, learning, and games*. Cambridge University Press, 2006.

[2] E. Camponogara, D. Jia, B. Krogh, and S. Talukdar, "Distributed model predictive control," *Control Systems, IEEE*, vol. 22, no. 1, pp. 44–52, February 2002.

[3] S. S. Ram, V. V. Veeravalli, and A. Nedić, "Distributed non-autonomous power control through distributed convex optimization," in *IEEE INFOCOM*, 2009, pp. 3001–3005.

[4] S. Kar and J. Moura, "Distributed consensus algorithms in sensor networks: Quantized data and random link failures," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1383 –1400, March 2010.

[5] A. Jadbabaie, J. Lin, and A. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor rules," *IEEE Transactions on Automatic Control*, vol. 48, no. 6, pp. 988 – 1001, June 2003.

[6] A. Nedić and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Transactions on Automatic Control*, vol. 54, no. 1, pp. 48–61, 2009.

[7] A. Nedić, A. Ozdaglar, and A. Parrilo, "Constrained consensus and optimization in multi-agent networks," *IEEE Transactions on Automatic Control*, vol. 55, no. 4, pp. 922–938, 2010.

[8] P. Bianchi and J. Jakubowicz, "Convergence of a multi-agent projected stochastic gradient algorithm," 2012, http://arxiv.org/abs/1107.2526.

[9] K. I. Tsianos, S. Lawlor, and M. G. Rabbat, "Consensus-based distributed optimization: Practical issues and applications in large-scale machine learning," in *Proceedings of the 50th Allerton Conference on Communication, Control, and Computing*, 2012.

[10] S. S. Ram, A. Nedić, and V. V. Veeravalli, "Asynchronous gossip algorithms for stochastic optimization," in *48th IEEE Conference on Decision and Control*, 2009, pp. 3581–3586.

[11] ——, "Asynchronous gossip algorithm for stochastic optimization: Constant stepsize analysis," in *Recent Advances in Optimization and its Applications in Engineering*, M. Diehl, F. Glineur, E. Jarlebring, and W. Michiels, Eds. Springer Berlin Heidelberg, 2010, pp. 51–60.

[12] A. Nedić, "Asynchronous broadcast-based convex optimization over a network," *IEEE Trans. Automat. Contr.*, vol. 56, no. 6, pp. 1337–1351, 2011.

[13] K. Srivastava and A. Nedić, "Distributed asynchronous constrained stochastic optimization," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 4, pp. 772–790, 2011.

[14] I. Lobel, A. Ozdaglar, and D. Feijer, "Distributed multi-agent optimization with state-dependent communication," *Mathematical Programming*, vol. 129, no. 2, pp. 255–284, 2011.

[15] 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.

[16] A. Nedić and A. Olshevsky, "Distributed optimization over time-varying directed graphs," *IEEE Transactions on Automatic Control*, vol. 60, no. 3, pp. 601–615, March 2015.

[17] K. Tsianos, S. Lawlor, and M. Rabbat, "Push-sum distributed dual-averaging for convex optimization," in *Proceedings of the 51st IEEE Conference on Decision and Control*, Maui, Hawaii, December 2012, pp. 5453–5458.

[18] J. Duchi, A. Agarwal, and M. 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] E. Wei and A. Ozdaglar, "On the $O(1/k)$ convergence of asynchronous distributed alternating direction method of multipliers," 2013, http://arxiv.org/pdf/1307.8254.

[20] F. Iutzeler, P. Bianchi, P. Ciblat, and W. Hachem, "Explicit convergence rate of a distributed alternating direction method of multipliers," 2013, http://arxiv.org/pdf/1312.1085.

[21] P. Bianchi, W. Hachem, and F. Iutzeler, "A stochastic coordinate descent primal-dual algorithm and applications to large-scale composite optimization," 2014, http://arxiv.org/pdf/1407.0898.

[22] F. Iutzeler, P. Bianchi, P. Ciblat, and W. Hachem, "Asynchronous distributed optimization using a randomized alternating direction method of multipliers," 2013, http://arxiv.org/pdf/1303.2837.

[23] W.Shi, Q. Ling, G. Wu, and W. Yin, "Extra: An exact first-order algorithm for decentralized consensus optimization," 2014, http://arxiv.org/pdf/1404.6264.

[24] A. Olshevsky and J. N. Tsitsiklis, "Convergence speed in distributed consensus and averaging," *SIAM J. Control Optim.*, vol. 48, no. 1, pp. 33–55, Feb. 2009.

[25] S. Lee and A. Nedić, "Distributed random projection algorithm for convex optimization," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, pp. 221–229, April 2013.

[26] B. Polyak, "Random algorithms for solving convex inequalities," in *Inherently Parallel Algorithms in Feasibility and Optimization and their Applications*, ser. Studies in Computational Mathematics, Y. C. Dan Butnariu and S. Reich, Eds. Elsevier, 2001, vol. 8, pp. 409 – 422.

[27] A. Nedić, "Random algorithms for convex minimization problems," *Mathematical Programming - B*, vol. 129, pp. 225–253, 2011.

[28] A. Nedić, "Random projection algorithms for convex set intersection problems," in *Proc. of the 49th IEEE Conference on Decision and Control*, 2010, pp. 7655–7660.

[29] T. Alamo, R. Tempo, and E. Camacho, "Randomized strategies for probabilistic solutions of uncertain feasibility and optimization problems," *IEEE Trans. Autom. Control*, vol. 54, no. 11, pp. 2545 –2559, Nov 2009.

[30] G. C. Calafiore, "Random convex programs," *SIAM J. Optimiz.*, vol. 20, no. 6, pp. 3427–3464, Dec. 2010.

[31] G. Calafiore and M. Campi, "Uncertain convex programs: Randomized solutions and confidence levels," *Mathematical Programming*, vol. 102, pp. 25–46, 2005.

[32] K. You and R. Tempo, "Networked parallel algorithms for robust convex optimization via the scenario approach," 2016, http://arxiv.org/abs/1607.05507.

[33] G. C. Calafiore and M. C. Campi, "Uncertain convex programs: randomized solutions and confidence levels," *Mathematical Programming*, vol. 102, pp. 25–46, 2004.

[34] ——, "The scenario approach to robust control design," *IEEE Transactions on Automatic Control*, vol. 51, no. 5, pp. 742–753, 2006.

[35] R. Tempo, G. Calafiore, and F. Dabbene, *Randomized Algorithms for Analysis and Control of Uncertain Systems, with Applications*. Springer-Verlag London, 2013.

[36] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*, ser. Studies in Applied Mathematics. Philadelphia, PA: SIAM, Jun. 1994, vol. 15.

[37] A. Nedić, A. Olshevsky, A. Ozdaglar, and J. Tsitsiklis, "Distributed subgradient methods and quantization effects," in *Proceedings of 47th IEEE Conference on Decision and Control*, December 2008, pp. 4177 –4184.

[38] B. Gharesifard and J. Cortes, "When does a digraph admit a doubly stochastic adjacency matrix?" in *Proceedings of the American Control Conference*, Baltimore, MD, 2010, pp. 2440–2445.

[39] A. Makhdoumi and A. Ozdaglar, "Graph balancing for distributed subgradient methods over directed graphs," in *54th IEEE Conference on Decision and Control*, 2015, pp. 1364–1371.

[40] L. Vandenberghe and S. P. Boyd, *Convex Optimization*. Cambridge University Press, 2004.

[41] B. Polyak, "Gradient methods for solving equations and inequalities," *USSR Comput. Math. and Math. Phys.*, vol. 4, no. 6, pp. 17–32, 1964.

[42] F. Facchinei and J.-S. Pang, *Finite-dimensional Variational Inequalities and Complementarity Problems*. Springer, New York, 2003.

[43] A. Lewis and J.-S. Pang, "Error bounds for convex inequality systems," in *Generalized Convexity*. Kluwer Academic Publishers, 1996, pp. 75–110.

[44] H. H. Bauschke and J. M. Borwein, "On projection algorithms for solving convex feasibility problems," *SIAM Rev.*, vol. 38, no. 3, pp. 367–426, Sep. 1996.

[45] J. V. Burke and M. C. Ferris, "Weak sharp minima in mathematical programming," *SIAM Journal on Control and Optimization*, vol. 31, pp. 1340–1359, 1993.

[46] F. Deutsch and H. Hundal, "The rate of convergence for the cyclic projections algorithm i: Angles between convex sets," *Journal of Approximation Theory*, vol. 142, no. 1, pp. 36 – 55, 2006.

[47] ——, "The rate of convergence for the cyclic projections algorithm ii: Norms of nonlinear operators," *Journal of Approximation Theory*, vol. 142, no. 1, pp. 56 – 82, 2006.

[48] ——, "The rate of convergence for the cyclic projections algorithm iii: Regularity of convex sets," *Journal of Approximation Theory*, vol. 155, no. 2, pp. 155 – 184, 2008.

[49] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra*. SIAM, 2000.

[50] D. P. Bertsekas, A. Nedić, and A. E. Ozdaglar, *Convex analysis and optimization*. Athena Scientific, 2003.

[51] B. Polyak, *Intro. to optimization*. Optimization software, Inc., Publications division, New York, 1987.

[52] S. S. Ram, A. Nedić, and V. V. Veeravalli, "A new class of distributed optimization algorithms: application to regression of distributed data," *Optimization Methods and Software*, vol. 27, no. 1, pp. 71–88, 2012.

[53] B. Polyak, "Minimization of unsmooth functionals," *USSR Computational Mathematics and Mathematical Physics*, vol. 9, no. 3, pp. 14 – 29, 1969.

[54] S. S. Ram, A. Nedić, and V. V. Veeravalli, "Distributed Stochastic Subgradient Projection Algorithms for Convex Optimization," *Journal of Optimization Theory and Applications*, vol. 147, pp. 516–545, 2010.

[55] V. D. Blondel, J. M. Hendrickx, A. Olshevsky, and J. N. Tsitsiklis, "Convergence in multiagent coordination, consensus, and flocking," 2005.

[56] L. Moreau, "Stability of multiagent systems with time-dependent communication links," *IEEE Transactions on Automatic Control*, vol. 50, no. 2, pp. 169–182, Feb 2005.

[57] A. Nedić and A. Ozdaglar, "Convergence rate for consensus with delays," *Journal of Global Optimization*, vol. 47, no. 3, pp. 437–456, 2010.

[58] J. Tsitsiklis, D. Bertsekas, and M. Athans, "Distributed asynchronous deterministic and stochastic gradient optimization algorithms," *IEEE Transactions on Automatic Control*, vol. 31, no. 9, pp. 803 – 812, Sep. 1986.

[59] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Trans. Inform. Theory*, vol. 52, no. 6, pp. 2508–2530, June 2006.

**Soomin Lee** is currently working as a Research Scientist at Yahoo! Labs. She received her Ph.D. in Electrical and Computer Engineering from the University of Illinois, Urbana-Champaign (2013). She received two master's degrees from the Korea Advanced Institute of Science and Technology in Electrical Engineering, and from the University of Illinois at Urbana-Champaign in Computer Science. After graduation, she worked as Postdoctoral Associate in Mechanical Engineering and Materials Science at Duke University and Postdoctoral Researcher in Industrial and Systems Engineering at Georgia Tech. Her research interests include theoretical optimization (convex, non-convex, online and stochastic), distributed control and optimization of various engineering systems interconnected over complex networks and large-scale machine learning in Internet industry.

**Michael M. Zavlanos** (S05M09) received the Diploma in mechanical engineering from the National Technical University of Athens (NTUA), Athens, Greece, in 2002, and the M.S.E. and Ph.D. degrees in electrical and systems engineering from the University of Pennsylvania, Philadelphia, PA, in 2005 and 2008, respectively.

He is currently an Assistant Professor in the Department of Mechanical Engineering and Materials Science at Duke University, Durham, NC. He also holds a secondary appointment in the Department of Electrical and Computer Engineering and the Department of Computer Science. Prior to joining Duke University, Dr. Zavlanos was an Assistant Professor in the Department of Mechanical Engineering at Stevens Institute of Technology, Hoboken, NJ, and a Postdoctoral Researcher in the GRASP Lab, University of Pennsylvania, Philadelphia, PA. His research interests include a wide range of topics in the emerging discipline of networked systems, with applications in robotic, sensor, and communication networks. He is particularly interested in hybrid solution techniques, on the interface of control theory, distributed optimization, estimation, and networking.

Dr. Zavlanos is a recipient of various awards including the 2014 Office of Naval Research Young Investigator Program (YIP) Award and the 2011 National Science Foundation Faculty Early Career Development (CAREER) Award.