Distributed Stochastic Subgradient Projection Algorithms for Convex Optimization

S. Sundhar Ram · A. Nedić · V.V. Veeravalli

Abstract We consider a distributed multi-agent network system where the goal is to minimize a sum of convex objective functions of the agents subject to a common convex constraint set. Each agent maintains an iterate sequence and communicates the iterates to its neighbors. Then, each agent combines weighted averages of the received iterates with its own iterate, and adjusts the iterate by using subgradient information (known with stochastic errors) of its own function and by projecting onto the constraint set.

The goal of this paper is to explore the effects of stochastic subgradient errors on the convergence of the algorithm. We first consider the behavior of the algorithm in mean, and then the convergence with probability 1 and in mean square. We consider general stochastic errors that have uniformly bounded second moments and obtain bounds on the limiting performance of the algorithm in mean for diminishing and non-diminishing stepsizes. When the means of the errors diminish, we prove that there is mean consensus between the agents and mean convergence to the optimum function value for diminishing stepsizes. When the mean errors diminish sufficiently fast, we strengthen the results to consensus and convergence of the iterates to an optimal solution with probability 1 and in mean square.

Keywords Distributed algorithm · Convex optimization · Subgradient methods · Stochastic approximation

S. Sundhar Ram · V.V. Veeravalli
Electrical and Computer Engineering Dept., University of Illinois at Urbana-Champaign, Urbana, IL, USA

A. Nedić (✉)
Industrial and Enterprise Systems Engineering Dept., University of Illinois at Urbana-Champaign, Urbana, IL, USA
e-mail: angelia@illinois.edu

1 Introduction

A number of problems that arise in the context of peer to peer networks can be posed as the constrained minimization of a sum of convex functions, where each component function is known only to one agent in the network [1–13]. In such networks it is not efficient, or possible, for an agent to share its objective function with other agents due to energy and privacy constraints. In addition, the network topology is often time varying due to agent mobility and link/node failures. Thus any algorithm used to solve the optimization problem in these networks must be distributed, i.e., each agent exchanges only limited information only with its immediate neighbors, and robust to changes in the network topology.

In literature two algorithms have been proposed for the above problem: *the Markov incremental subgradient algorithm* [6, 10] and *the distributed subgradient algorithm* studied in [4] (and its later version [7]). The key difference between these two algorithms is in the mechanism that is used to distribute the computations. In the Markov incremental algorithm, the agents sequentially update a single iterate sequence by passing the iterate to each other. When an agent receives the iterate, it updates the iterate using its own objective function and then passes the iterate to a randomly chosen neighbor. Thus, the order in which the agents update the iterate follows a Markov chain (time inhomogeneous due to network topology changes) making the algorithm stochastic by design. In addition, a single iterate sequence is generated. In contrast, the distributed subgradient algorithm is deterministic by design and generates multiple iterate sequences (one sequence per agent). In the distributed subgradient algorithms of [4, 7], each agent maintains an iterate sequence and communicates the iterates to its neighbors. Then, each agent averages the received iterates with its own iterate (similar to a consensus step), and updates the average using its own function.

In this paper, we study the *effect of stochastic errors on the distributed subgradient algorithm*, where the errors are due to the agents computing the noisy subgradients of their respective objective functions.[1] We first consider general stochastic errors and stepsizes, and obtain error bounds on the limiting performance of the algorithm. We then consider errors whose mean and stepsize diminish sufficiently fast, and prove consensus and convergence of the iterates to an optimal solution both with probability 1 and in mean square. The extension of the analysis for the distributed subgradient from the deterministic case in [7] to the case with stochastic subgradient errors is non-trivial. This is primarily because the stochastic subgradient errors made by each agent propagate through the network to every other agent and also across time, making the iterates statistically dependent across time and agents. Further, the constraint introduces a non-linearity to the problem.

There are significant differences in the behavior of the algorithm of this paper and the algorithm of [6], where the effect of stochastic errors on the Markov incremental algorithm is studied. In the distributed algorithm, each agent maintains and updates a different iterate sequence. This gives rise to a deterministic consensus error (disagreement between agents) that interacts with the stochastic subgradient errors. In

---

[1]This is motivated by many applications including distributed learning and recursive regression over networks.

the Markov incremental algorithm the random nature of the algorithm (each agent passes the iterate to a random neighbor) results in a stochastic process that interacts with the stochastic subgradient errors. The difference in the nature of the interacting errors produces different convergence behavior of these algorithms. Inherently, their analysis is also significantly different which is reflected in the scope of the results obtained. We summarize the main results obtained in this paper and compare it with the corresponding results in [6] for the Markov incremental algorithm.

- Theorem 5.3 obtains bounds on the performance of the distributed subgradient algorithm after a finite number of iterations for diminishing and non-diminishing stepsizes and general stochastic errors. No such results are available for the Markov incremental algorithm in [6].
- Theorem 6.2 obtains sufficient conditions on the stepsizes and the stochastic errors for the iterates converge to an optimal point with probability 1 and in mean square. Theorem 4.3 of [6] obtains only mean square convergence of the Markov incremental algorithm for a smaller class of stepsizes and only for the case when the constraint set is bounded.
- Theorem 5.2 obtain bounds on the *lim inf* of the expected function value at the iterates generated by the distributed subgradient algorithm for diminishing and non-diminishing stepsizes and general stochastic errors. This bound is of the order of m^2, where m is the total number of agents in the network. The corresponding bound for the Markov incremental algorithm in Theorem 4.4 of [6] is of the order of m.
- Theorem 5.3 establishes the *lim sup* bound for the expected function value along the time-averaged sequence of the distributed algorithm with diminishing and non-diminishing stepsizes, and general stochastic errors. No such bound is available for the Markov incremental algorithm in [6].

We next discuss related literature. The distributed subgradient algorithm was studied with constant stepsizes for the unconstrained and error free case in [4]. As mentioned the paper that is most related is [7]. In this a general algorithm is proposed for the case when even the constraint set is not common across the agents and each agent has a unique different set. While the problem and the proposed algorithm are general, the analysis is performed only for two special cases (a) when the network is completely connected, i.e., each agent shares its iterate with every other agent in each iteration, and (b) when that constraint set at all the agents are the same (the case studied in this paper). However, the paper considers only diminishing stepsize and there are no stochastic errors in the subgradient terms.

Distributed incremental algorithms have been studied in [1, 2, 6, 9, 10, 14]. The effects of stochastic errors on these algorithms have been investigated in [6, 14–19]. As mentioned, in the incremental algorithms all the agents update a single iterate sequence and are hence fundamentally different from the distributed subgradient algorithm in which each agent updates a different iterate sequence. Also related are the optimization algorithms in [20–22]. However, these algorithms are not local as the complete objective function information is available to each and every agent, with the aim of distributing the processing. The work in this paper is also related at a much broader level to the distributed consensus algorithms [7, 20–29]. In these algorithms,

each agent starts with a different value and through local information exchange, the agents eventually agree on a common value. The effect of random errors on consensus algorithms have been investigated in [25–27, 30]. In addition, since we are interested in the effect of stochastic errors, our paper is also related to the literature on stochastic gradient methods [31–33].

The rest of the paper is organized as follows. In Sect. 2, we formulate the problem, describe the algorithm and state our basic assumptions. In Sect. 3 we state some results from literature that we use in the analysis, while in Sect. 4, we derive two important lemmas that form the backbone of the analysis. In Sect. 5, we study the convergence properties of the method in mean, and in Sect. 6 we focus on the convergence properties with probability 1 and in mean square. Finally, we provide some concluding remarks in Sect. 7.

2 Problem, Algorithm and Assumptions

In this section, we formulate the problem of interest and describe the algorithm that we propose. We also state and discuss our assumptions on the agent connectivity and information exchange.

2.1 Problem

We consider a network of m agents that are indexed by $1, \ldots, m$. Often, when convenient, we index the agents by using set $V = \{1, \ldots, m\}$. The network objective is to solve the following constrained optimization problem:

$$
\begin{aligned}
\text{minimize} \quad & \sum_{i=1}^{m} f_i(x) \\
\text{subject to} \quad & x \in X,
\end{aligned}
\tag{1}
$$

where $X \subseteq \Re^n$ is a constraint set and $f_i : X \to \Re$ for all i. Related to the problem, we use the following notation

$$
f(x) = \sum_{i=1}^{m} f_i(x), \qquad f^* = \min_{x \in X} f(x), \qquad X^* = \{x \in X : f(x) = f^*\}.
$$

We are interested in the case when the problem in (2) is convex. Specifically, we assume that the following assumption holds.

Assumption 2.1 *The functions f_i and the set X are such that*

(a) *The set X is closed and convex.*
(b) *The functions f_i, $i \in V$ are defined and convex over an open set that contains the set X.*

The function f_i is known only partially to agent i in the sense that the agent can only obtain a noisy estimate of the function subgradient. The goal is to solve problem (2) using an algorithm that is distributed and local.[2]

We make no assumption on the differentiability of the functions f_i. At points where the gradient does not exist, we use the notion of subgradients. A vector ∇f_i is a subgradient of f_i at a point $x \in \text{dom } f$ if the following relation holds

$$\nabla f_i(x)^T(y - x) \le f_i(y) - f_i(x) \quad \text{for all } y \in \text{dom } f. \tag{2}$$

Since the set X is contained in an open set over which the functions are defined and convex, a subgradient of f_i exists at any point of the set X (see [35] or [36]).

2.2 Algorithm

To solve the problem in (2) with its inherent decentralized information access, we consider an iterative subgradient method. The iterations are distributed accordingly among the agents, whereby each agent i is minimizing its convex objective f_i over the set X and locally exchanging the iterates with its neighbors.

Let $w_{i,k}$ be the iterate with agent i at the end of iteration k. At the beginning of iteration $k + 1$, agent i receives the current iterate of a subset of the agents. Then, agent i computes a weighted average of these iterates and adjusts this average along the negative subgradient direction of f_i, which is computed with stochastic errors. The adjusted iterate is then projected onto the constraint set X. Formally, each agent i generates its iterate sequence $\{w_{i,k}\}$ according to the following relation:

$$w_{i,k+1} = P_X[v_{i,k} - \alpha_{k+1}(\nabla f_i(v_{i,k}) + \epsilon_{i,k+1})], \tag{3}$$

starting with some initial iterate $w_{i,0} \in X$. Here, $\nabla f_i(v_{i,k})$ denotes the subgradient of f_i at $v_{i,k}$ and $\epsilon_{i,k+1}$ is the stochastic error in the subgradient evaluation. The scalar $\alpha_{k+1} > 0$ is the stepsize and P_X denotes the Euclidean projection onto the set X. The vector $v_{i,k}$ is the weighted average computed by agent i and is given by

$$v_{i,k} = \sum_{j \in N_i(k+1)} a_{i,j}(k+1) w_{j,k}, \tag{4}$$

where $N_i(k + 1)$ denotes the set of agents whose current iterates are available to agent i in the $(k + 1)$-st iteration. We assume that $i \in N_i(k + 1)$ for all agents and at all times k. The scalars $a_{i,j}(k + 1)$ are the non-negative weights that agent i assigns to agent j's iterate. We will find it convenient to define $a_{i,j}(k+1)$ as 0 for $j \notin N_i(k+1)$ and rewrite (4) as

$$v_{i,k} = \sum_{j=1}^{m} a_{i,j}(k+1) w_{j,k}. \tag{5}$$

This is a "consensus"-based step ensuring that, in a long run, the information of each f_i reaches every agent with the same frequency, directly or through a sequence of

---

[2]See [6, 34] for wireless network applications that can be cast in this framework.

local communications. Due to this, the iterates $w_{j,k}$ become eventually "the same" for all j and for large enough k. The update step in (3) is just a subgradient iteration for minimizing f_i over X taken after the "consensus"-based step.

2.3 Additional Assumptions

In addition to Assumption 2.1, we make some assumptions on the inter-agent exchange model and the weights. The first assumption requires the agents to communicate sufficiently often so that all the component functions, directly or indirectly, influence the iterate sequence of any agent. Recall that we defined $N_i(k+1)$ as the set of agents that agent i communicates with in iteration $k+1$. Define (V, E_{k+1}) to be the graph with edges

$$E_{k+1} = \{(j, i) : j \in N_i(k+1), i \in V\}.$$

Assumption 2.2 *There exists a scalar Q such that the graph $(V, \bigcup_{l=1,\ldots,Q} E_{k+l})$ is strongly connected for all k.*

It is also important that the influence of the functions f_i is "equal" in a long run so that the sum of the component functions is minimized rather than a weighted sum of them. The influence of a component f_j on the iterates of agent i depends on the weights that agent i uses. To ensure equal influence, we make the following assumption on the weights.

Assumption 2.3 *For $i \in V$ and all k,*

(a) $a_{i,j}(k+1) \geq 0$, *and* $a_{i,j}(k+1) = 0$ *when* $j \notin N_i(k+1)$,
(b) $\sum_{j=1}^{m} a_{i,j}(k+1) = 1$,
(c) *There exists a scalar $\eta \in (0, 1)$ such that $a_{i,j}(k+1) \geq \eta$ when $j \in N_i(k+1)$,*
(d) $\sum_{i=1}^{m} a_{i,j}(k+1) = 1$.

Assumptions 2.3a and 2.3b state that each agent calculates a weighted average of all the iterates it has access to. Assumption 2.3c ensures that each agent gives a sufficient weight to its current iterate and all the iterates it receives.[3] Assumption 2.3d, together with Assumption 2.2, as we will see later, ensures that all the agents are equally influential in the long run. In other words, Assumption 2.3d is crucial to ensure that $\sum_{i=1}^{m} f_i$ is minimized as opposed to a weighted sum of the functions f_i with non-equal weights. To satisfy Assumption 2.3d, the agents need to coordinate their weights. Some coordination schemes are discussed in [4, 6].

---

[3]The agents need not be aware of the common bound η.

3 Preliminaries

In this section, we state some results for future reference.

Euclidean norm inequalities. For any vectors $v_1, \ldots, v_M \in \Re^n$, we have

$$\sum_{i=1}^{M} \left\| v_i - \frac{1}{M} \sum_{j=1}^{M} v_j \right\|^2 \leq \sum_{i=1}^{M} \| v_i - x \|^2 \quad \text{for any } x \in \Re^n. \tag{6}$$

The preceding relation states that the average of a finite set of vectors minimizes the sum of distances between each vector and any vector in $\Re^n$, which can be verified using the first-order optimality conditions.

Both the Euclidean norm and its square are convex functions, i.e., for any vectors $v_1, \ldots, v_M \in \Re^n$ and nonnegative scalars $\beta_1, \ldots, \beta_M$ such that $\sum_{i=1}^{M} \beta_i = 1$, we have

$$\left\| \sum_{i=1}^{M} \beta_i v_i \right\| \leq \sum_{i=1}^{M} \beta_i \| v_i \|, \tag{7}$$

$$\left\| \sum_{i=1}^{M} \beta_i v_i \right\|^2 \leq \sum_{i=1}^{M} \beta_i \| v_i \|^2. \tag{8}$$

The following inequality is the well-known[4] *non-expansive* property of the Euclidean projection onto a nonempty, closed and convex set X,

$$\| P_X[x] - P_X[y] \| \leq \| x - y \| \quad \text{for all } x, y \in \Re^n. \tag{9}$$

Scalar sequences. We make use of the following result. Its proof is in Appendix.

Lemma 3.1 *Let $\{\gamma_k\}$ be a scalar sequence.*

(a) *If $\lim_{k \to \infty} \gamma_k = \gamma$ and $0 < \beta < 1$, then $\lim_{k \to \infty} \sum_{\ell=0}^{k} \beta^{k-\ell} \gamma_\ell = \frac{\gamma}{1-\beta}$.*

(b) *If $\gamma_k \geq 0$ for all k, $\sum_k \gamma_k < \infty$ and $0 < \beta < 1$, then $\sum_{k=0}^{\infty} (\sum_{\ell=0}^{k} \beta^{k-\ell} \gamma_\ell) < \infty$.*

(c) *If $\limsup_{k \to \infty} \gamma_k = \gamma$ and $\{\zeta_k\}$ is a positive scalar sequence with $\sum_{k=1}^{\infty} \zeta_k = \infty$, then $\limsup_{K \to \infty} \frac{\sum_{k=0}^{K} \gamma_k \zeta_k}{\sum_{k=0}^{K} \zeta_k} \leq \gamma$. In addition, if $\liminf_{k \to \infty} \gamma_k = \gamma$, then $\lim_{K \to \infty} \frac{\sum_{k=0}^{K} \gamma_k \zeta_k}{\sum_{k=0}^{K} \zeta_k} = \gamma$.*

Matrix convergence. Let $A(k)$ be the matrix with (i, j)-th entry equal to $a_{i,j}(k)$. As a consequence of Assumptions 2.3a, 2.3b and 2.3d, the matrix $A(k)$ is doubly stochastic.[5] Define

$$\Phi(k, s) = A(k) A(k-1) \cdots A(s+1) \quad \text{for all } k, s \text{ with } k \geq s \geq 0. \tag{10}$$

---

[4] See for example [35], Proposition 2.2.1.

[5] The sum of its entries in every row and in every column is equal to 1.

We next state a result from [29] (Corollary 1) on the convergence properties of the matrix $\Phi(k, s)$. Let $[\Phi(k, s)]_{i,j}$ denote the (i, j)-th entry of the matrix $\Phi(k, s)$, and let $e \in \Re^m$ be the column vector with all entries equal to 1.

Lemma 3.2 *Let Assumptions* 2.2 *and* 2.3 *hold. Then*

(a) $\lim_{k \to \infty} \Phi(k, s) = \frac{1}{m} e e^T$ *for all* $s \geq 0$.
(b) $|[\Phi(k, s)]_{i,j} - \frac{1}{m}| \leq \theta \beta^{k-s}$ *for all* $k \geq s \geq 0$, *where* $\theta = (1 - \frac{\eta}{4m^2})^{-2}$ *and* $\beta = (1 - \frac{\eta}{4m^2})^{\frac{1}{Q}}$.

Stochastic convergence. The following result is due to Robbins and Siegmund (Lemma 11, Chap. 2.2, [37]).

Theorem 3.1 *Let* $\{B_k\}$, $\{D_k\}$, *and* $\{H_k\}$ *be non-negative random sequences and let* $\{\zeta_k\}$ *be a deterministic nonnegative scalar sequence. Let* G_k *be the* σ-*algebra generated by* $B_1, \ldots, B_k, D_1, \ldots, D_k, H_1, \ldots, H_k$. *Suppose that* $\sum_k \zeta_k < \infty$,

$$\mathsf{E}[B_{k+1} \mid G_k] \leq (1 + \zeta_k) B_k - D_k + H_k \quad \text{for all } k,$$

and $\sum_k H_k < \infty$ *with probability* 1. *Then, the sequence* $\{B_k\}$ *converges to a non-negative random variable and* $\sum_k D_k < \infty$ *with probability* 1, *and in mean.*

We also use the well known Fatou's lemma [38], stating that for a sequence of non-negative random variables $\{X_k\}$, we have

$$\mathsf{E}\left[\liminf_{k \to \infty} X_k\right] \leq \liminf_{k \to \infty} \mathsf{E}[X_k].$$

4 Basic Relations

In this section, we derive two basic relations that form the basis for the analysis in this paper. The first of them deals with the disagreements among the agents, and the second deals with the agent iterate sequences.

4.1 Disagreement Estimate

The agent disagreements are typically thought of as the norms $\|w_{i,k} - w_{j,k}\|$ of the differences between the iterates $w_{i,k}$ and $w_{j,k}$ generated by different agents according to (3)–(4). Alternatively, the agent disagreements can be measured with respect to a reference sequence, which we adopt here. In particular, we study the behavior of $\|y_k - w_{i,k}\|$, where y_k is the vector defined by

$$y_k = \frac{1}{m} \sum_{i=1}^m w_{i,k} \quad \text{for all } k. \tag{11}$$

In the next lemma, we provide a basic estimate for $\|y_k - w_{j,k}\|$. The rate of convergence result from Lemma 3.2 plays a crucial role in obtaining this estimate.

Lemma 4.1 *Let Assumptions* 2.1a, 2.2, *and* 2.3 *hold. Assume that the subgradients of f_i are uniformly bounded over the set X, i.e., there are scalars C_i such that*

$$\|\nabla f_i(x)\| \le C_i \quad \text{for all } x \in X \text{ and all } i \in V.$$

Then, for all $j \in V$ and $k \ge 0$,

$$\|y_{k+1} - w_{j,k+1}\| \le m\theta\beta^{k+1} \max_{i \in V} \|w_{i,0}\| + \theta \sum_{\ell=1}^{k} \alpha_\ell \beta^{k+1-\ell} \sum_{i=1}^{m} (C_i + \|\epsilon_{i,\ell}\|)$$

$$+ \frac{\alpha_{k+1}}{m} \sum_{i=1}^{m} (C_i + \|\epsilon_{i,k+1}\|) + \alpha_{k+1}(C_j + \|\epsilon_{j,k+1}\|).$$

Proof Define for all $i \in V$ and all k,

$$p_{i,k+1} = w_{i,k+1} - \sum_{j=1}^{m} a_{i,j}(k+1)w_{j,k}. \tag{12}$$

Using the matrices $\Phi(k,s)$ defined in (10) we can write

$$w_{j,k+1} = \sum_{i=1}^{m} [\Phi(k+1,0)]_{j,i} w_{i,0} + p_{j,k+1} + \sum_{\ell=1}^{k} \left(\sum_{i=1}^{m} [\Phi(k+1,\ell)]_{j,i} p_{i,\ell} \right). \tag{13}$$

Using (12), we can also rewrite y_k, defined in (11), as follows

$$y_{k+1} = \frac{1}{m} \left(\sum_{i=1}^{m} \sum_{j=1}^{m} a_{i,j}(k+1) w_{j,k} + \sum_{i=1}^{m} p_{i,k+1} \right)$$

$$= \frac{1}{m} \left(\sum_{j=1}^{m} \left(\sum_{i=1}^{m} a_{i,j}(k+1) \right) w_{j,k} + \sum_{i=1}^{m} p_{i,k+1} \right).$$

In the view of the doubly stochasticity of the weights, we have $\sum_{i=1}^{m} a_{i,j}(k+1) = 1$, implying that

$$y_{k+1} = \frac{1}{m} \left(\sum_{j=1}^{m} w_{j,k} + \sum_{i=1}^{m} p_{i,k+1} \right) = y_k + \frac{1}{m} \sum_{i=1}^{m} p_{i,k+1}.$$

Therefore

$$y_{k+1} = y_0 + \frac{1}{m} \sum_{\ell=1}^{k+1} \sum_{i=1}^{m} p_{i,\ell} = \frac{1}{m} \sum_{i=1}^{m} w_{i,0} + \frac{1}{m} \sum_{\ell=1}^{k+1} \sum_{i=1}^{m} p_{i,\ell}. \tag{14}$$

Substituting for y_{k+1} from (14) and for $w_{j,k+1}$ from (13), we obtain

$$
\|y_{k+1} - w_{j,k+1}\| = \left\| \frac{1}{m} \sum_{i=1}^{m} w_{i,0} + \frac{1}{m} \sum_{\ell=1}^{k+1} \sum_{i=1}^{m} p_{i,\ell} \right.
$$

$$
- \left(\sum_{i=1}^{m} [\Phi(k+1,0)]_{j,i} \, w_{i,0} + p_{j,k+1} \right.
$$

$$
\left. \left. + \sum_{\ell=1}^{k} \sum_{i=1}^{m} [\Phi(k+1,\ell)]_{j,i} \, p_{i,\ell} \right) \right\|
$$

$$
= \left\| \sum_{i=1}^{m} \left(\frac{1}{m} - [\Phi(k+1,0)]_{j,i} \right) w_{i,0} \right.
$$

$$
+ \sum_{\ell=1}^{k} \sum_{i=1}^{m} \left(\frac{1}{m} - [\Phi(k+1,\ell)]_{j,i} \right) p_{i,\ell}
$$

$$
\left. + \left(\frac{1}{m} \sum_{i=1}^{m} p_{i,k+1} - p_{j,k+1} \right) \right\|.
$$

Therefore, for all $j \in V$ and all k,

$$
\|y_{k+1} - w_{j,k+1}\| \le \sum_{i=1}^{m} \left| \frac{1}{m} - [\Phi(k+1,0)]_{j,i} \right| \|w_{i,0}\|
$$

$$
+ \sum_{\ell=1}^{k} \sum_{i=1}^{m} \left| \frac{1}{m} - [\Phi(k+1,\ell)]_{j,i} \right| \|p_{i,\ell}\|
$$

$$
+ \frac{1}{m} \sum_{i=1}^{m} \|p_{i,k+1}\| + \|p_{j,k+1}\|.
$$

We can bound $\|w_{i,0}\| \le \max_{i \in V} \|w_{i,0}\|$. Further, we can use the rate of convergence result from Lemma 3.2 to bound $\left| \frac{1}{m} - [\Phi(k,\ell)]_{j,i} \right|$. We obtain

$$
\|y_{k+1} - w_{j,k+1}\| \le m \theta \beta^{k+1} \max_{i \in V} \|w_{i,0}\| + \theta \sum_{\ell=1}^{k} \beta^{k+1-\ell} \sum_{i=1}^{m} \|p_{i,\ell}\|
$$

$$
+ \frac{1}{m} \sum_{i=1}^{m} \|p_{i,k+1}\| + \|p_{j,k+1}\|. \tag{15}
$$

We next estimate the norms of the vectors $\|p_{i,k}\|$ for any k. From the definition of $p_{i,k+1}$ in (12) and the definition of the vector $v_{i,k}$ in (4), we have $p_{i,k+1} = w_{i,k+1} - v_{i,k}$. Note that, being a convex combination of vectors $w_{j,k}$ in the convex set X, the

vector $v_{i,k}$ is in the set X. By the definition of the iterate $w_{i,k+1}$ in (3) and the non-expansive property of the Euclidean projection in (9), we have

$$\|p_{i,k+1}\| = \|P_X[v_{i,k} - \alpha_{k+1}(\nabla f_i(v_{i,k}) + \epsilon_{i,k+1})] - v_{i,k}\|$$
$$\leq \alpha_{k+1}\|\nabla f_i(v_{i,k}) + \epsilon_{i,k+1}\|$$
$$\leq \alpha_{k+1}(C_i + \|\epsilon_{i,k+1}\|).$$

In the last step we have used the subgradient boundedness. By substituting the preceding relation in (15), we obtain the desired relation. $\qquad\square$

4.2 Iterate Relation

Here, we derive a relation for the distances $\|v_{i,k+1} - z\|$ and the function value differences $f(y_k) - f(z)$ for an arbitrary $z \in X$. This relation together with Lemma 4.1 provides the basis for our subsequent convergence analysis. In what follows, recall that $f = \sum_{i=1}^m f_i$.

Lemma 4.2 *Let Assumptions* 2.1, 2.2, *and* 2.3 *hold. Assume that the subgradients of* f_i *are uniformly bounded over the set* X, *i.e., there are scalars* C_i *such that*

$$\|\nabla f_i(x)\| \leq C_i \quad \text{for all } x \in X \text{ and all } i \in V.$$

Then, for any $z \in X$ *and all* k,

$$\sum_{i=1}^m \|v_{i,k+1} - z\|^2 \leq \sum_{i=1}^m \|v_{i,k} - z\|^2 - 2\alpha_{k+1}(f(y_k) - f(z))$$
$$+ 2\alpha_{k+1}\left(\max_{i \in V} C_i\right) \sum_{j=1}^m \|y_k - w_{j,k}\|$$
$$- 2\alpha_{k+1} \sum_{i=1}^m \epsilon_{i,k+1}^T(v_{i,k} - z) + \alpha_{k+1}^2 \sum_{i=1}^m (C_i + \|\epsilon_{i,k+1}\|)^2.$$

Proof Using the Euclidean projection property in (9), from the definition of the iterate $w_{i,k+1}$ in (3), we have for any $z \in X$ and all k,

$$\|w_{i,k+1} - z\|^2 = \|P_X[v_{i,k} - \alpha_{k+1}(\nabla f_i(v_{i,k}) + \epsilon_{i,k+1})] - z\|^2$$
$$\leq \|v_{i,k} - z\|^2 - 2\alpha_{k+1}\nabla f_i(v_{i,k})^T(v_{i,k} - z) - 2\alpha_{k+1}\epsilon_{i,k+1}^T(v_{i,k} - z)$$
$$+ \alpha_{k+1}^2 \|\nabla f_i(v_{i,k}) + \epsilon_{i,k+1}\|^2.$$

By using the subgradient inequality in (2) to bound the second term, we obtain

$$\|w_{i,k+1} - z\|^2 \leq \|v_{i,k} - z\|^2 - 2\alpha_{k+1}(f_i(v_{i,k}) - f_i(z))$$
$$- 2\alpha_{k+1}\epsilon_{i,k+1}^T(v_{i,k} - z) + \alpha_{k+1}^2 \|\nabla f_i(v_{i,k}) + \epsilon_{i,k+1}\|^2. \quad (16)$$

Note that by the convexity of the squared norm [cf. (8)], we have

$$\sum_{i=1}^{m} \|v_{i,k+1} - z\|^2 = \sum_{i=1}^{m} \left\| \sum_{j=1}^{m} a_{i,j}(k+2)w_{j,k+1} - z \right\|^2$$

$$\leq \sum_{i=1}^{m} \sum_{j=1}^{m} a_{i,j}(k+2)\|w_{j,k+1} - z\|^2.$$

In view of Assumption 2.3, we have $\sum_{i=1}^{m} a_{i,j}(k+2) = 1$ for all j and k, implying that

$$\sum_{i=1}^{m} \|v_{i,k+1} - z\|^2 \leq \sum_{j=1}^{m} \|w_{j,k+1} - z\|^2.$$

By summing the relations in (16) over all $i \in V$ and by using the preceding relation, we obtain

$$\sum_{i=1}^{m} \|v_{i,k+1} - z\|^2$$

$$\leq \sum_{i=1}^{m} \|v_{i,k} - z\|^2 - 2\alpha_{k+1} \sum_{i=1}^{m} (f_i(v_{i,k}) - f_i(z))$$

$$- 2\alpha_{k+1} \sum_{i=1}^{m} \epsilon_{i,k+1}^T (v_{i,k} - z) + \alpha_{k+1}^2 \sum_{i=1}^{m} \|\nabla f_i(v_{i,k}) + \epsilon_{i,k+1}\|^2. \qquad (17)$$

From (2) we have

$$f_i(v_{i,k}) - f_i(z) \geq (f_i(v_{i,k}) - f_i(y_k)) + (f_i(y_k) - f_i(z))$$

$$\geq -\|\nabla f_i(v_{i,k})\| \|y_k - v_{i,k}\| + (f_i(y_k) - f_i(z)). \qquad (18)$$

Recall that $v_{i,k} = \sum_{j=1}^{m} a_{i,j}(k+1)w_{j,k}$ [cf. (5)]. Substituting for $v_{i,k}$ and using the convexity of the norm [cf. (7)] and $f + \sum_{i=1}^{m} f_i$, from (18) we obtain

$$\sum_{i=1}^{m} f_i(v_{i,k}) - f_i(z) \geq -\sum_{i=1}^{m} \|\nabla f_i(v_{i,k})\| \left\| y_k - \sum_{j=1}^{m} a_{i,j}(k+1)w_{j,k} \right\|$$

$$+ (f(y_k) - f(z))$$

$$\geq -\sum_{i=1}^{m} \|\nabla f_i(v_{i,k})\| \sum_{j=1}^{m} a_{i,j}(k+1)\|y_k - w_{j,k}\|$$

$$+ (f(y_k) - f(z))$$

$$\geq -\left(\max_{i \in V} \|\nabla f_i(v_{i,k})\|\right) \sum_{j=1}^{m} \left(\sum_{i=1}^{m} a_{i,j}(k+1)\right) \|y_k - w_{j,k}\|$$

$$+ (f(y_k) - f(z))$$

$$= -\Big(\max_{i \in V} \|\nabla f_i(v_{i,k})\|\Big) \sum_{j=1}^{m} \|y_k - w_{j,k}\| + (f(y_k) - f(z)).$$

By using the preceding estimate in relation (17), we have

$$\sum_{i=1}^{m} \|v_{i,k+1} - z\|^2 \leq \sum_{i=1}^{m} \|v_{i,k} - z\|^2 - 2\alpha_{k+1}(f(y_k) - f(z))$$

$$+ 2\alpha_{k+1}\Big(\max_{i \in V} \|\nabla f_i(v_{i,k})\|\Big) \sum_{j=1}^{m} \|y_k - w_{j,k}\|$$

$$- 2\alpha_{k+1} \sum_{i=1}^{m} \epsilon_{i,k+1}^T (v_{i,k} - z) + \alpha_{k+1}^2 \sum_{i=1}^{m} \|\nabla f_i(v_{i,k}) + \epsilon_{i,k+1}\|^2.$$

The result follows by using the subgradient norm boundedness, $\|\nabla f_i(v_{i,k})\| \leq C_i$ for all k and i.   □

5 Convergence in Mean

Here, we study the behavior of the iterates generated by the algorithm, under the assumption that the errors have bounded norms in mean square. In particular, we assume the following.

Assumption 5.1 *The subgradient errors are uniformly bounded in mean square, i.e., there are scalars $\bar{v}_i$ such that*

$$\mathsf{E}\big[\|\epsilon_{i,k+1}\|^2\big] \leq \bar{v}_i^2 \quad \textit{for all } i \in V \textit{ and all } k.$$

Using this assumption, we provide a bound on the expected disagreement $\mathsf{E}[\|w_{i,k} - y_k\|]$ for nondiminishing stepsize. We later use this bound to provide an estimate for the algorithm's performance in mean. The bound is provided in the following theorem.

Theorem 5.1 *Let Assumptions 2.1a, 2.2, 2.3 and 5.1 hold. Also, let the subgradients of each f_i be uniformly bounded over X, i.e., for each $i \in V$ there is C_i such that*

$$\|\nabla f_i(x)\| \leq C_i \quad \textit{for all } x \in X.$$

If the stepsize $\{\alpha_k\}$ is such that $\lim_{k \to \infty} \alpha_k = \alpha$ for some $\alpha \geq 0$, then for all $j \in V$,

$$\limsup_{k \to \infty} \mathsf{E}\,[\|y_{k+1} - w_{j,k+1}\|] \leq \alpha \max_{i \in V}\{C_i + \bar{v}_i\}\Big(2 + \frac{m\theta\beta}{1 - \beta}\Big).$$

Proof The conditions of Lemma 4.1 are satisfied. Taking the expectation in the relation of Lemma 4.1 and using the inequality $\mathsf{E}[\|\epsilon_{i,k}\|] \le \sqrt{\mathsf{E}[\|\epsilon_{i,k}\|^2]} = \bar{\nu}_i$, we obtain for all $j \in V$ and all k,

$$\mathsf{E}[\|y_{k+1} - w_{j,k+1}\|] \le m\theta\beta^{k+1} \max_{i \in V} \|w_{i,0}\| + m\theta\beta \max_{i \in V}\{C_i + \bar{\nu}_i\} \sum_{\ell=1}^{k} \beta^{k-\ell}\alpha_\ell$$

$$+ 2\alpha_{k+1} \max_{i \in V}\{C_i + \bar{\nu}_i\}.$$

Since $\lim_{k\to\infty} \alpha_k = \alpha$, by Lemma 3.1(a) we have $\lim_{k\to\infty} \sum_{\ell=1}^{k} \beta^{k-\ell}\alpha_\ell = \frac{\alpha}{1-\beta}$. Using this equality and $\lim_{k\to\infty} \alpha_k = \alpha$, we obtain the result by letting $k \to \infty$ in the preceding relation.   □

When the stepsize is diminishing (i.e., $\alpha = 0$), the result of Theorem 5.1 implies that the expected disagreements $\mathsf{E}[\|y_{k+1} - w_{j,k+1}\|]$ converge to 0 for all j. Thus, there is an asymptotic consensus in mean. We formally state this as a corollary.

Corollary 5.1 *Let the conditions of Theorem 5.1 hold with $\alpha = 0$. Then $\lim_{k\to\infty} \mathsf{E}[\|w_{j,k} - y_k\|] = 0$ for all $j \in V$.*

We next obtain bounds on the performance of the algorithm. We make the additional assumption that the set X is bounded. Thus, the subgradients of each f_i are also bounded (see [35], Proposition 4.2.3).

Note that, under Assumption 5.1, by Jensen's inequality we have $\|\mathsf{E}[\epsilon_{i,k+1}]\| \le \bar{\nu}_i$. Therefore, under Assumption 5.1,

$$\limsup_{k\to\infty} \|\mathsf{E}[\epsilon_{i,k+1}]\| \le \bar{\nu}_i \quad \text{for all } i \in V. \tag{19}$$

We have used this relation in our analysis of the agent disagreements in Theorem 5.1. Using this relation, we obtain special results for the cases when the errors are zero mean or when their mean is diminishing, i.e., the cases $\mathsf{E}[\epsilon_{i,k+1}] = 0$ for all i, k, or $\limsup_{k\to\infty} \|\mathsf{E}[\epsilon_{i,k+1}]\| = 0$ for all i.

Theorem 5.2 *Let Assumptions 2.1, 2.2, 2.3 and 5.1 hold. Assume that the set X is bounded. Let $\lim_{k\to\infty} \alpha_k = \alpha$ with $\alpha \ge 0$. If $\alpha = 0$, also assume that $\sum_k \alpha_k = \infty$. Then, for all $j \in V$,*

$$\liminf_{k\to\infty} \mathsf{E}[f(w_{j,k})] \le f^* + \max_{x,y \in X} \|x - y\| \sum_{i=1}^{m} \bar{\mu}_i + m\alpha \left(\max_{i \in V}\{C_i + \bar{\nu}_i\}\right)^2$$

$$\times \left(\frac{9}{2} + \frac{2m\theta\beta}{1-\beta}\right),$$

where $\bar{\mu}_i = \limsup_{k\to\infty} \|\mathsf{E}[\epsilon_{i,k+1}]\|$ and C_i is an upper-bound on the subgradient norms of f_i over the set X.

Proof Under Assumption 5.1, the limit superiors $\bar{\mu}_i = \limsup_{k \to \infty} \|\mathsf{E}[\epsilon_{i,k+1}]\|$ are finite [cf. (19)]. Since the set X is bounded the subgradients of f_i over the set X are also bounded for each $i \in V$; hence, the bounds C_i, $i \in V$ on subgradient norms exist. Thus, the conditions of Lemma 4.2 are satisfied. Further, by Assumption 2.1, the set X is contained in the interior of the domain of f, over which the function is continuous (by convexity; see [36]). Thus, the set X is compact and f is continuous over X, implying that the optimal set X^* is nonempty. Let $x^* \in X^*$, and let $y = x^*$ in Lemma 4.2. We have, for all k,

$$
\sum_{i=1}^{m} \|v_{i,k+1} - x^*\|^2 \leq \sum_{i=1}^{m} \|v_{i,k} - x^*\|^2 - 2\alpha_{k+1}(f(y_k) - f^*)
$$

$$
+ 2\alpha_{k+1}\left(\max_{i \in V} C_i\right) \sum_{j=1}^{m} \|y_k - w_{j,k}\|
$$

$$
- 2\alpha_{k+1} \sum_{i=1}^{m} \epsilon_{i,k+1}^T (v_{i,k} - x^*) + \alpha_{k+1}^2 \sum_{i=1}^{m} (C_i + \|\epsilon_{i,k+1}\|)^2.
$$

Since X is bounded, by using $\|v_{i,k} - x^*\| \leq \max_{x,y \in X} \|x - y\|$, taking the expectation and using the error bounds $\mathsf{E}[\|\epsilon_{i,k+1}\|^2] \leq \bar{v}_i^2$ we obtain

$$
\sum_{i=1}^{m} \mathsf{E}[\|v_{i,k+1} - x^*\|^2] \leq \sum_{i=1}^{m} \mathsf{E}[\|v_{i,k} - x^*\|^2] - 2\alpha_{k+1}(\mathsf{E}[f(y_k)] - f^*)
$$

$$
+ 2\alpha_{k+1}\left(\max_{i \in V} C_i\right) \sum_{j=1}^{m} \mathsf{E}[\|y_k - w_{j,k}\|]
$$

$$
+ 2\alpha_{k+1} \max_{x,y \in X} \|x - y\| \sum_{i=1}^{m} \|\mathsf{E}[\epsilon_{i,k+1}]\|
$$

$$
+ \alpha_{k+1}^2 \sum_{i=1}^{m} (C_i + \bar{v}_i)^2. \tag{20}
$$

By rearranging the terms and summing over $k = 1, \ldots, K$, for an arbitrary K, we obtain

$$
2 \sum_{k=1}^{K} \alpha_{k+1}\left((\mathsf{E}[f(y_k)] - f^*) - \left(\max_{i \in V} C_i\right) \sum_{j=1}^{m} \mathsf{E}[\|y_k - w_{j,k}\|] \right.
$$

$$
\left. - \max_{x,y \in X} \|x - y\| \sum_{i=1}^{m} \|\mathsf{E}[\epsilon_{i,k+1}]\| - \frac{m\alpha_{k+1}}{2}\left(\max_{i \in V}\{C_i + \bar{v}_i\}\right)^2 \right)
$$

$$
\leq \sum_{i=1}^{m} \mathsf{E}[\|v_{i,1} - x^*\|^2] - \sum_{i=1}^{m} \mathsf{E}[\|v_{i,K+1} - x^*\|^2] \leq m \max_{x,y \in X} \|x - y\|^2.
$$

Note that when $\alpha_{k+1} \to \alpha$ and $\alpha > 0$, we have $\sum_k \alpha_k = \infty$. When $\alpha = 0$, we have assumed that $\sum_k \alpha_k = \infty$. Therefore, by letting $K \to \infty$, we have

$$\liminf_{k \to \infty} \left(\mathsf{E}\,[f(y_k)] - \left(\max_{i \in V} C_i \right) \sum_{j=1}^{m} \mathsf{E}\,[\|y_k - w_{j,k}\|] \right.$$
$$\left. - \max_{x,y \in X} \|x - y\| \sum_{i=1}^{m} \|\mathsf{E}\,[\epsilon_{i,k+1}]\| - \frac{m\alpha_{k+1}}{2} \left(\max_{i \in V} \{C_i + \bar{\nu}_i\} \right)^2 \right) \le f^*.$$

Using $\limsup_{k \to \infty} \|\mathsf{E}\,[\epsilon_{i,k+1}]\| = \bar{\mu}_i$ [see (19)] and $\lim_{k \to \infty} \alpha_k = \alpha$, we obtain

$$\liminf_{k \to \infty} \mathsf{E}\,[f(y_k)] \le f^* + \frac{m\alpha}{2} \left(\max_{i \in V} \{C_i + \bar{\nu}_i\} \right)^2$$
$$+ \left(\max_{i \in V} C_i \right) \sum_{j=1}^{m} \limsup_{k \to \infty} \mathsf{E}\,[\|y_k - w_{j,k}\|]$$
$$+ \max_{x,y \in X} \|x - y\| \sum_{i=1}^{m} \bar{\mu}_i.$$

Next from the convexity inequality in (2) and the boundedness of the subgradients it follows that for all k and $j \in V$,

$$\mathsf{E}\,[f(w_{j,k}) - f(y_k)] \le \left(\sum_{i=1}^{m} C_i \right) \mathsf{E}\,[\|y_k - w_{j,k}\|],$$

implying

$$\liminf_{k \to \infty} \mathsf{E}\,[f(w_{j,k})] \le f^* + \frac{m\alpha}{2} \left(\max_{i \in V} \{C_i + \bar{\nu}_i\} \right)^2$$
$$+ \left(\max_{i \in V} C_i \right) \sum_{j=1}^{m} \limsup_{k \to \infty} \mathsf{E}\,[\|y_k - w_{j,k}\|]$$
$$+ \left(\sum_{i=1}^{m} C_i \right) \limsup_{k \to \infty} \mathsf{E}\,[\|y_k - w_{j,k}\|] + \max_{x,y \in X} \|x - y\| \sum_{i=1}^{m} \bar{\mu}_i.$$

By Theorem 5.1, we have for all $j \in V$,

$$\limsup_{k \to \infty} \mathsf{E}\,[\|y_k - w_{j,k}\|] \le \alpha \max_{i \in V} \{C_i + \bar{\nu}_i\} \left(2 + \frac{m\theta\beta}{1 - \beta} \right).$$

By using the preceding relation, we see that

$$\liminf_{k \to \infty} \mathsf{E}\,[f(w_{j,k})] \le f^* + \frac{m\alpha}{2} \left(\max_{i \in V} \{C_i + \bar{\nu}_i\} \right)^2 + \max_{x,y \in X} \|x - y\| \sum_{i=1}^{m} \bar{\mu}_i$$

$$+ m\alpha \left(\max_{i \in V} C_i \right) \max_{i \in V} \{C_i + \bar{v}_i\} \left(2 + \frac{m\theta\beta}{1 - \beta} \right)$$

$$+ \alpha \left(\sum_{i=1}^{m} C_i \right) \max_{j \in V} \{C_j + \bar{v}_j\} \left(2 + \frac{m\theta\beta}{1 - \beta} \right)$$

$$\leq f^* + \max_{x,y \in X} \|x - y\| \sum_{i=1}^{m} \bar{\mu}_i + m\alpha \left(\max_{i \in V} \{C_i + \bar{v}_i\} \right)^2$$

$$\times \left(\frac{9}{2} + \frac{2m\theta\beta}{1 - \beta} \right). \qquad \qquad \square$$

The network topology influences the error only through the term $\frac{\theta\beta}{1-\beta}$ and can hence be used as a figure of merit for comparing different topologies. For a network that is strongly connected at every time, [i.e., $Q = 1$ in Assumption 2.2] and when η in Assumption 2.3 does not depend on the number m of agents, the term $\frac{\theta\beta}{1-\beta}$ is of the order m^2 and the error bound scales as m^4.

We next show that stronger bounds can be obtained for a specific weighted time averages of the iterates $w_{i,k}$. In particular, we investigate the limiting behavior of $\{f(z_{i,t})\}$, where $z_{i,t} = \frac{\sum_{k=1}^{t} \alpha_{k+1} w_{i,k}}{\sum_{k=1}^{t} \alpha_{k+1}}$. Note that agent i can locally and recursively evaluate $z_{i,t+1}$ from $z_{i,t}$ and $w_{i,t+1}$.

Theorem 5.3 *Consider the weighted time averages $z_{j,t} = \frac{\sum_{k=1}^{t} \alpha_{k+1} w_{j,k}}{\sum_{k=1}^{t} \alpha_{k+1}}$ for $j \in V$ and $t \geq 1$. Let the conditions of Theorem 5.2 hold. Then, we have for all $j \in V$,*

$$\limsup_{t \to \infty} \mathsf{E}\,[f(z_{j,t})] \leq f^* + \max_{x,y \in X} \|x - y\| \sum_{i=1}^{m} \bar{\mu}_i + m\alpha \left(\max_{i \in V} \{C_i + \bar{v}_i\} \right)^2 \left(\frac{9}{2} + \frac{2m\theta\beta}{1 - \beta} \right).$$

Proof The relation in (20) of Theorem 5.2 is valid, and we have for any $x^* \in X^*$,

$$\sum_{i=1}^{m} \mathsf{E}\,[\|v_{i,k+1} - x^*\|^2] \leq \sum_{i=1}^{m} \mathsf{E}\,[\|v_{i,k} - x^*\|^2] - 2\alpha_{k+1}(\mathsf{E}\,[f(y_k)] - f^*)$$

$$+ 2\alpha_{k+1} \left(\max_{i \in V} C_i \right) \sum_{\ell=1}^{m} \mathsf{E}\,[\|y_k - w_{\ell,k}\|]$$

$$+ 2\alpha_{k+1} \max_{x,y \in X} \|x - y\| \sum_{i=1}^{m} \|\mathsf{E}\,[\epsilon_{i,k+1}]\|$$

$$+ \alpha_{k+1}^2 \sum_{i=1}^{m} (C_i + \bar{v}_i)^2.$$

From the subgradient boundedness and the subgradient inequality in (2) we have for any j,

$$\mathsf{E}[f(y_k)] - \mathsf{E}[f(w_{j,k})] \geq -\left(\sum_{i=1}^{m} C_i\right) \mathsf{E}[\|y_k - w_{j,k}\|]$$

$$\geq -m\left(\max_{i \in V} C_i\right) \mathsf{E}[\|y_k - w_{j,k}\|].$$

Therefore, we obtain

$$\sum_{i=1}^{m} \mathsf{E}[\|v_{i,k+1} - x^*\|^2]$$

$$\leq \sum_{i=1}^{m} \mathsf{E}[\|v_{i,k} - x^*\|^2] - 2\alpha_{k+1}(\mathsf{E}[f(w_{j,k})] - f^*)$$

$$+ 2\alpha_{k+1}\left(\max_{i \in V} C_i\right)\left(m\mathsf{E}[\|y_k - w_{j,k}\|] + \sum_{i=1}^{m} \mathsf{E}[\|y_k - w_{i,k}\|]\right)$$

$$+ 2\alpha_{k+1} \max_{x,y \in X} \|x - y\| \sum_{i=1}^{m} \|\mathsf{E}[\epsilon_{i,k+1}]\| + \alpha_{k+1}^2 \sum_{i=1}^{m} (C_i + \bar{v}_i)^2.$$

By re-arranging these terms, summing over $k = 1, \ldots, t$ and dividing with $2\sum_{k=1}^{t} \alpha_{k+1}$, we further obtain

$$\sum_{k=1}^{t} \frac{\alpha_{k+1} \mathsf{E}[f(w_{j,k})]}{\sum_{k=1}^{t} \alpha_{k+1}}$$

$$\leq f^* + \frac{1}{2\sum_{k=1}^{t} \alpha_{k+1}} \sum_{i=1}^{m} \mathsf{E}[\|v_{i,1} - x^*\|^2]$$

$$+ \sum_{k=1}^{t} \frac{\alpha_{k+1}(\max_{i \in V} C_i)(m\mathsf{E}[\|y_k - w_{j,k}\|] + \sum_{i=1}^{m} \mathsf{E}[\|y_k - w_{i,k}\|])}{\sum_{k=1}^{t} \alpha_{k+1}}$$

$$+ \max_{x,y \in X} \|x - y\| \sum_{i=1}^{m} \frac{\sum_{k=1}^{t} \alpha_{k+1} \|\mathsf{E}[\epsilon_{i,k+1}]\|}{\sum_{k=1}^{t} \alpha_{k+1}} + \frac{\sum_{k=1}^{t} \alpha_{k+1}^2}{2\sum_{k=1}^{t} \alpha_{k+1}} \sum_{i=1}^{m} (C_i + \bar{v}_i)^2.$$

Next by the convexity of f note that

$$f(z_{j,t}) = f\left(\sum_{k=1}^{t} \frac{\alpha_{k+1} w_{j,k}}{\sum_{k=1}^{t} \alpha_{k+1}}\right) \leq \sum_{k=1}^{t} \frac{\alpha_{k+1} f(w_{j,k})}{\sum_{k=1}^{t} \alpha_{k+1}}.$$

From the preceding two relations we obtain

$$\mathsf{E}\,[f(z_{j,t})]$$

$$\leq f^* + \frac{1}{2\sum_{k=1}^{t}\alpha_{k+1}} \sum_{i=1}^{m} \mathsf{E}\left[\|v_{i,1} - x^*\|^2\right]$$

$$+ \sum_{k=1}^{t} \frac{\alpha_{k+1}(\max_{i\in V} C_i)(m\mathsf{E}\,[\|y_k - w_{j,k}\|] + \sum_{i=1}^{m}\mathsf{E}\,[\|y_k - w_{i,k}\|])}{\sum_{k=1}^{t}\alpha_{k+1}}$$

$$+ \max_{x,y\in X}\|x - y\| \sum_{i=1}^{m} \frac{\sum_{k=1}^{t}\alpha_{k+1}\|\mathsf{E}\,[\epsilon_{i,k+1}]\|}{\sum_{k=1}^{t}\alpha_{k+1}} + \frac{\sum_{k=1}^{t}\alpha_{k+1}^2}{2\sum_{k=1}^{t}\alpha_{k+1}} \sum_{i=1}^{m}(C_i + \bar{v}_i)^2.$$

(21)

First note that in the limit as $t \to \infty$, the second term in (21) converges to 0 since $\sum_{k=1}^{t}\alpha_{k+1} = \infty$. By using the results of Lemma 3.1c, for the remaining terms, we obtain

$$\limsup_{t\to\infty}\mathsf{E}\,[f(z_{j,t})]$$

$$\leq f^* + \left(\max_{i\in V} C_i\right)\limsup_{k\to\infty}\left(m\mathsf{E}\,[\|y_k - w_{j,k}\|] + \sum_{i=1}^{m}\mathsf{E}\,[\|y_k - w_{i,k}\|]\right)$$

$$+ \max_{x,y\in X}\|x - y\| \sum_{i=1}^{m}\limsup_{k\to\infty}\|\mathsf{E}\,[\epsilon_{i,k+1}]\| + \frac{\alpha}{2}\sum_{i=1}^{m}(C_i + \bar{v}_i)^2.$$

By Theorem 5.1, we have for all $j \in V$,

$$\limsup_{k\to\infty}\mathsf{E}\,[\|y_k - w_{j,k}\|] \leq \alpha \max_{i\in V}\{C_i + \bar{v}_i\}\left(2 + \frac{m\theta\beta}{1 - \beta}\right),$$

which when substituted in the preceding relation, yields

$$\limsup_{t\to\infty}\mathsf{E}\,[f(z_{j,t})] \leq f^* + 2m\alpha\left(\max_{i\in V} C_i\right)\max_{i\in V}\{C_i + \bar{v}_i\}\left(2 + \frac{m\theta\beta}{1 - \beta}\right)$$

$$+ \max_{x,y\in X}\|x - y\| \sum_{i=1}^{m}\limsup_{k\to\infty}\|\mathsf{E}\,[\epsilon_{i,k+1}]\| + \frac{\alpha}{2}\sum_{i=1}^{m}(C_i + \bar{v}_i)^2$$

$$\leq f^* + \max_{x,y\in X}\|x - y\| \sum_{i=1}^{m}\limsup_{k\to\infty}\|\mathsf{E}\,[\epsilon_{i,k+1}]\|$$

$$+ m\alpha\left(\max_{i\in V}\{C_i + \bar{v}_i\}\right)^2\left(\frac{9}{2} + \frac{2m\theta\beta}{1 - \beta}\right). \qquad \square$$

The error bounds in Theorems 5.2 and 5.3 have the same form, but they apply to different sequences of function evaluations. Furthermore, in Theorem 5.3, the bound is for *all* subsequences of $\mathsf{E}[f(z_{i,k})]$ for each agent i. In contrast, in Theorem 5.2, the bound is only for *a* subsequence of $\mathsf{E}[f(z_{i,k})]$ for each agent i. Theorem 5.3 demonstrates that, due to the convexity of the objective function f, there is an advantage when agents are using the running averages of their iterates.

When the error[6] moments $\|\mathsf{E}[\epsilon_{i,k+1}]\|$ converge to zero as $k \to \infty$, and the stepsize converges to zero $[\alpha = 0]$, Theorems 5.2 and 5.3 yield respectively

$$\liminf_{k\to\infty} \mathsf{E}[f(w_{j,k})] = f^* \quad \text{and} \quad \lim_{k\to\infty} \mathsf{E}[f(z_{j,k})] = f^*.$$

When a constant stepsize α is used, the vector $z_{j,t}$ is simply the running average of all the iterates of agent j until time t, i.e., $z_{j,t} = \frac{1}{t}\sum_{k=1}^{t} w_{j,k}$. For this case, with zero mean errors, the relation in (21) reduces to

$$
\mathsf{E}[f(z_{j,t})] \le f^* + \frac{1}{2t\alpha}\sum_{i=1}^{m} \mathsf{E}[\|v_{i,1} - x^*\|^2]
$$

$$
+ \left(\max_{i\in V} C_i\right)\frac{1}{t}\sum_{k=1}^{t}\left(m\mathsf{E}[\|y_k - w_{j,k}\|] + \sum_{i=1}^{m}\mathsf{E}[\|y_k - w_{i,k}\|]\right)
$$

$$
+ \frac{\alpha}{2}\sum_{i=1}^{m}(C_i + \bar{v}_i)^2. \tag{22}
$$

This can be used to derive an estimate per iteration, as seen in the following.

Corollary 5.2 *Under the conditions of Theorem 5.2 with $\|\mathsf{E}[\epsilon_{i,k+1}]\| = 0$ and $\alpha_k = 0$ for all i and k, for the average sequences $\{z_{j,k}\}$ we have for all t and j,*

$$
\mathsf{E}[f(z_{j,t})] \le f^* + \frac{1}{2t\alpha}\sum_{i=1}^{m}\mathsf{E}\left[\|v_{i,1} - x^*\|^2\right] + \frac{2m^2\theta\beta^2}{t(1-\beta)}\left(\max_{i\in V} C_i\right)\left(\max_{i\in V}\|w_{i,0}\|\right)
$$

$$
+ m\alpha\left(\max_{i\in V}\{C_i + \bar{v}_i\}\right)^2\left(\frac{9}{2} + \frac{2m\theta\beta}{1-\beta}\right).
$$

Proof Taking the expectation in the relation of Lemma 4.1, we obtain

$$
\mathsf{E}[\|y_{k+1} - w_{j,k+1}\|] \le m\theta\beta^{k+1}\max_{i\in V}\|w_{i,0}\| + m\alpha\theta\beta\left(\max_{i\in V}\{C_i + \bar{v}_i\}\right)\sum_{\ell=1}^{k}\beta^{k-\ell}
$$

$$
+ 2\alpha\max_{i\in V}\{C_i + \bar{v}_i\}
$$

$$
\le m\theta\beta^{k+1}\max_{i\in V}\|w_{i,0}\| + \alpha\left(\max_{i\in V}\{C_i + \bar{v}_i\}\right)\left(2 + \frac{m\theta\beta}{1-\beta}\right).
$$

---

[6]When the moments $\|\mathsf{E}[\epsilon_{i,k+1}]\|$ are zero, it can be seen that the results of Theorems 5.2 and 5.3 hold when the boundedness of X is replaced by the weaker assumption that the subgradients of each f_i are bounded over X.

Combining the preceding relation with the inequality in (22), and using $\sum_{k=1}^{t} \beta^{k+1} \leq \frac{\beta^2}{1-\beta}$, we obtain

$$\mathsf{E}[f(z_{j,t})] \leq f^* + \frac{1}{2t\alpha} \sum_{i=1}^{m} \mathsf{E}\left[\|v_{i,1} - x^*\|^2\right] + \frac{2m^2\theta\beta^2}{t(1-\beta)} \left(\max_{i \in V} C_i\right) \left(\max_{i \in V} \|w_{i,0}\|\right)$$

$$+ m\alpha \left(\max_{i \in V}\{C_i + \bar{v}_i\}\right)^2 \left(\frac{9}{2} + \frac{2m\theta\beta}{1-\beta}\right). \qquad \Box$$

The preceding equation provides a bound on the algorithm's performance at each iteration. The bound can be used in obtaining stopping rules for the algorithm. For example, consider the error free case ($\bar{v}_i = 0$) and suppose that the goal is to determine the number of iterations required for agents to find a point in the ϵ-optimal set, i.e., in the set $X_\epsilon = \{x \in X : f(x) \leq f^* + \epsilon\}$. Minimizing the bound in Corollary 5.2 over different stepsize values α, we can show that ϵ-optimality can be achieved in $N_\epsilon = \lceil \frac{1}{\psi_\epsilon^2} \rceil$ iterations with a stepsize $\alpha_\epsilon = \frac{\sqrt{A}\psi_\epsilon}{\sqrt{C}}$, where ψ_ϵ is the positive root of the quadratic equation

$$Bx^2 + 2\sqrt{AC}x - \epsilon = 0,$$

and A, B and C are

$$A = \frac{1}{2} \sum_{i=1}^{m} \|v_{i,1} - x^*\|^2, \qquad B = \frac{2m^2\theta\beta^2}{1-\beta} \left(\max_{i \in V} C_i\right) \left(\max_{i \in V} \|w_{i,0}\|\right),$$

$$C = m \left(\max_{i \in V}\{C_i + \bar{v}_i\}\right)^2 \left(\frac{9}{2} + \frac{2m\theta\beta}{1-\beta}\right).$$

Since ψ_ϵ scales as $\sqrt{\epsilon}$, we can conclude that N_ϵ scales as $\frac{1}{\epsilon^2}$. Equivalently, we can say that the level ϵ of sub-optimality diminishes inversely with the square root of the number of iterations.

6 Almost Sure and Mean Square Convergence

In this section, we impose some additional assumptions on the subgradient errors to obtain almost sure consensus among the agents and almost sure convergence of the iterates to an optimal solution of (2). Towards this, define F_k to be the σ-algebra $\sigma(\epsilon_{i,\ell};\ i \in V, 0 \leq \ell \leq k)$ generated by the errors in the agent system up to time k. In other words, F_k captures the history of the errors until the end of time k. We use the following assumption on the subgradient errors $\epsilon_{i,k}$.

Assumption 6.1 *There are scalars v_i such that $\mathsf{E}[\|\epsilon_{i,k+1}\|^2 \mid F_k] \leq v_i^2$ for all k with probability 1.*

Note that Assumption 6.1 is stronger than Assumption 5.1. Furthermore, when the errors are independent across iterations and across agents, Assumption 6.1 reduces to Assumption 5.1.

We start by analyzing the agents' disagreements measured in terms of distances $\|y_k - w_{j,k}\|$. We have the following result.

Theorem 6.1 *Let Assumptions* 2.1a, 2.2, 2.3 *and* 6.1 *hold. Suppose that the subgradients of each* f_i *are uniformly bounded over* X, *i.e., for each* $i \in V$ *there is* C_i *such that*

$$\|\nabla f_i(x)\| \le C_i \quad \text{for all } x \in X.$$

If $\sum_{k=0}^{\infty} \alpha_{k+1}^2 < \infty$, *then with probability* 1,

$$\sum_{k=1}^{\infty} \alpha_{k+2} \|y_{k+1} - w_{j,k+1}\| < \infty \quad \text{for all } j \in V.$$

Furthermore, for all $j \in V$, *we have* $\lim_{k \to \infty} \|y_{k+1} - w_{j,k+1}\| = 0$ *with probability* 1 *and in mean square.*

Proof By Lemma 4.1 and the subgradient boundedness, we have for all $j \in V$,

$$\|y_{k+1} - w_{j,k+1}\| \le m\theta\beta^{k+1} \max_{i \in V} \|w_{i,0}\| + \theta \sum_{\ell=1}^{k} \beta^{k+1-\ell} \sum_{i=1}^{m} \alpha_\ell (C_i + \|\epsilon_{i,\ell}\|)$$

$$+ \frac{1}{m} \sum_{i=1}^{m} \alpha_{k+1} (C_i + \|\epsilon_{i,k+1}\|) + \alpha_{k+1} (C_j + \|\epsilon_{j,k+1}\|).$$

Using the inequalities

$$\alpha_{k+2}\alpha_\ell (C_i + \|\epsilon_{i,\ell}\|) \le \frac{1}{2} \big(\alpha_{k+2}^2 + \alpha_\ell^2 (C_i + \|\epsilon_{i,\ell}\|)^2\big)$$

and $(C_i + \|\epsilon_{i,\ell}\|)^2 \le 2C_i^2 + 2\|\epsilon_{i,\ell}\|^2$, we obtain

$$\alpha_{k+2} \|y_{k+1} - w_{j,k+1}\| \le \alpha_{k+2} m\theta\beta^{k+1} \max_{i \in V} \|w_{i,0}\|$$

$$+ \theta \sum_{\ell=1}^{k} \beta^{k+1-\ell} \sum_{i=1}^{m} \left(\frac{1}{2}\alpha_{k+2}^2 + \alpha_\ell^2 \big(C_i^2 + \|\epsilon_{i,\ell}\|^2\big)\right)$$

$$+ \frac{1}{m} \sum_{i=1}^{m} \left(\frac{1}{2}\alpha_{k+2}^2 + \alpha_{k+1}^2 \big(C_i^2 + \|\epsilon_{i,k+1}\|^2\big)\right)$$

$$+ \frac{1}{2}\alpha_{k+2}^2 + \alpha_{k+1}^2 \big(C_j^2 + \|\epsilon_{j,k+1}\|^2\big).$$

By using the inequalities $\sum_{\ell=1}^{k} \beta^{k+1-\ell} \le \frac{\beta}{1-\beta}$ for all $k \ge 1$ and $\frac{1}{2m} + \frac{1}{2} \le 1$, and by grouping the terms accordingly, from the preceding relation we have

$$
\alpha_{k+2}\|y_{k+1} - w_{j,k+1}\| \le \alpha_{k+2} m\theta\beta^{k+1} \max_{i \in V} \|w_{i,0}\| + \left(1 + \frac{m\theta\beta}{2(1-\beta)}\right)\alpha_{k+2}^2
$$

$$
+ \theta \sum_{\ell=1}^{k} \alpha_{\ell}^2 \beta^{k+1-\ell} \sum_{i=1}^{m}(C_i^2 + \|\epsilon_{i,\ell}\|^2)
$$

$$
+ \frac{1}{m}\alpha_{k+1}^2 \sum_{i=1}^{m}(C_i^2 + \|\epsilon_{i,k+1}\|^2) + \alpha_{k+1}^2(C_j^2 + \|\epsilon_{j,k+1}\|^2).
$$

Taking the conditional expectation and using $\mathsf{E}[\|\epsilon_{i,\ell}\|^2 \mid F_{\ell-1}] \le v_i^2$, and then taking the expectation again, we obtain

$$
\mathsf{E}[\alpha_{k+2}\|y_{k+1} - w_{j,k+1}\|] \le \alpha_{k+2} m\theta\beta^{k+1} \max_{i \in V} \|w_{i,0}\| + \left(1 + \frac{m\theta\beta}{2(1-\beta)}\right)\alpha_{k+2}^2
$$

$$
+ \theta \left(\sum_{i=1}^{m}(C_i^2 + v_i^2)\right) \sum_{\ell=1}^{k} \alpha_{\ell}^2 \beta^{k+1-\ell}
$$

$$
+ \frac{1}{m}\alpha_{k+1}^2 \sum_{i=1}^{m}(C_i^2 + v_i^2) + \alpha_{k+1}^2(C_j^2 + v_j^2).
$$

Since $\sum_k \alpha_k^2 < \infty$ (and hence $\{\alpha_k\}$ bounded), the first two terms and the last two terms are summable. Furthermore, in view of Lemma 3.1 [part (b)], we have $\sum_{k=1}^{\infty} \sum_{\ell=1}^{k} \beta^{k+1-\ell}\alpha_{\ell}^2 < \infty$. Thus, the third term is also summable. Hence $\sum_{k=1}^{\infty} \mathsf{E}[\alpha_{k+2}\|y_{k+1} - w_{j,k+1}\|] < \infty$. From the monotone convergence theorem [38], it follows that

$$
\mathsf{E}\left[\sum_{k=1}^{\infty}\alpha_{k+2}\|y_{k+1} - w_{j,k+1}\|\right] = \sum_{k=1}^{\infty}\mathsf{E}[\alpha_{k+2}\|y_{k+1} - w_{j,k+1}\|],
$$

and it is hence finite for all j. If the expected value of a random variable is finite, then the variable has to be finite with probability 1; thus, with probability 1,

$$
\sum_{k=1}^{\infty}\alpha_{k+2}\|y_{k+1} - w_{j,k+1}\| < \infty \quad \text{for all } j \in V. \tag{23}
$$

We now show that $\lim_{k\to\infty}\|y_k - w_{j,k}\| = 0$ with probability 1 for all $j \in V$. Note that the conditions of Theorem 5.1 are satisfied with $\bar{v}_i = v_i$ and $\alpha = 0$. Therefore, $\|y_k - w_{j,k}\|$ converges to 0 in the mean and from Fatou's lemma it follows that

$$
0 \le \mathsf{E}\left[\liminf_{k\to\infty}\|y_k - w_{j,k}\|\right] \le \liminf_{k\to\infty}\mathsf{E}[\|y_k - w_{j,k}\|] = 0,
$$

and hence $\mathsf{E}\left[\liminf_{k\to\infty}\|y_k - w_{j,k}\|\right] = 0$. Therefore, with probability 1,

$$\liminf_{k\to\infty}\|y_k - w_{j,k}\| = 0. \tag{24}$$

To complete the proof, in view of (24) it suffices to show that $\|y_k - w_{j,k}\|$ converges with probability 1. To show this, we define

$$r_{i,k+1} = \sum_{j=1}^{m} a_{i,j}(k+1)w_j(k) - \alpha_{k+1}(\nabla f_i(v_{i,k}) + \epsilon_{i,k+1}),$$

and note that $P_X[r_{i,k+1}] = w_{i,k+1}$ [see (3) and (4)]. Since $y_k = \frac{1}{m}\sum_{i=1}^{m} w_{i,k}$ and the set X is convex, it follows that $y_k \in X$ for all k. Therefore, by the non-expansive property of the Euclidean projection in (9), we have $\|w_{i,k+1} - y_k\|^2 \le \|r_{i,k+1} - y_k\|^2$ for all $i \in V$ and all k. Summing these relations over all i, we obtain

$$\sum_{i=1}^{m}\|w_{i,k+1} - y_k\|^2 \le \sum_{i=1}^{m}\|r_{i,k+1} - y_k\|^2 \quad \text{for all } k.$$

From $y_{k+1} = \frac{1}{m}\sum_{i=1}^{m} w_{i,k+1}$ and the fact that the average of vectors minimizes the sum of distances between each vector and arbitrary vector in $\Re^n$ [cf. (6)], we further obtain

$$\sum_{i=1}^{m}\|w_{i,k+1} - y_{k+1}\|^2 \le \sum_{i=1}^{m}\|w_{i,k+1} - y_k\|^2.$$

Therefore, for all k,

$$\sum_{i=1}^{m}\|w_{i,k+1} - y_{k+1}\|^2 \le \sum_{i=1}^{m}\|r_{i,k+1} - y_k\|^2. \tag{25}$$

We next relate $\sum_{i=1}^{m}\|r_{i,k+1} - y_k\|^2$ to $\sum_{i=1}^{m}\|w_{i,k} - y_k\|^2$. From the definition of $r_{i,k+1}$ and the equality $\sum_{j=1}^{m} a_{i,j}(k+1) = 1$ [cf. Assumption 2.3b], we have

$$r_{i,k+1} - y_k = \sum_{j=1}^{m} a_{i,j}(k+1)(w_{j,k} - y_k) - \alpha_{k+1}(\nabla f_i(v_{i,k}) + \epsilon_{i,k+1}).$$

By Assumption 2.3a and 2.3b, we have that the weights $a_{i,j}(k+1)$, $j \in V$ yield a convex combination. Thus, by the convexity of the norm [(7) and (8)] and by the subgradient boundedness, we have

$$\|r_{i,k+1} - y_k\|^2 \le \sum_{j=1}^{m} a_{i,j}(k+1)\|w_{j,k} - y_k\|^2 + \alpha_{k+1}^2\|\nabla f_i(v_{i,k}) + \epsilon_{i,k+1}\|^2$$

$$+ 2\alpha_{k+1}\|\nabla f_i(v_{i,k}) + \epsilon_{i,k+1}\|\sum_{j=1}^{m} a_{i,j}(k+1)\|w_{j,k} - y_k\|$$

$$\leq \sum_{j=1}^{m} a_{i,j}(k+1)\|w_{j,k} - y_k\|^2 + 2\alpha_{k+1}^2 \left(C_i^2 + \|\epsilon_{i,k+1}\|^2\right)$$

$$+ 2\alpha_{k+1}(C_i + \|\epsilon_{i,k+1}\|) \sum_{j=1}^{m} a_{i,j}(k+1)\|w_{j,k} - y_k\|.$$

Summing over all i and using $\sum_{i=1}^{m} a_{i,j}(k+1) = 1$ [cf. Assumption 2.3d], we obtain

$$\sum_{i=1}^{m} \|r_{i,k+1} - y_k\|^2 \leq \sum_{j=1}^{m} \|w_{j,k} - y_k\|^2 + 2\alpha_{k+1}^2 \sum_{i=1}^{m} \left(C_i^2 + \|\epsilon_{i,k+1}\|^2\right)$$

$$+ 2\alpha_{k+1} \sum_{i=1}^{m} (C_i + \|\epsilon_{i,k+1}\|) \sum_{j=1}^{m} a_{i,j}(k+1)\|w_{j,k} - y_k\|.$$

Using this in (25) and taking the conditional expectation, we see that for all k, we have with probability 1,

$$\sum_{i=1}^{m} \mathsf{E}\left[\|w_{i,k+1} - y_{k+1}\|^2 \mid F_k\right] \leq \sum_{i=1}^{m} \|w_{i,k} - y_k\|^2 + 2\alpha_{k+1}^2 \sum_{i=1}^{m} \left(C_i^2 + v_i^2\right)$$

$$+ 2\alpha_{k+1} \sum_{i=1}^{m} (C_i + v_i) \sum_{j=1}^{m} \|w_{j,k} - y_k\|, \quad (26)$$

where we use $a_{i,j}(k+1) \leq 1$ for all i, j and k, and the relations $\mathsf{E}[\|\epsilon_{i,k+1}\|^2 \mid F_k] \leq v_i^2$, $\mathsf{E}[\|\epsilon_{i,k+1}\| \mid F_k] \leq v_i$ holding with probability 1.

We now apply Theorem 3.1 to the relation in (26). To verify that the conditions of Theorem 3.1 are satisfied, note that the stepsize satisfies $\sum_{k=1}^{\infty} \alpha_{k+1}^2 < \infty$ for all $i \in V$. We also have $\sum_{k=1}^{\infty} \alpha_{k+1}\|w_{j,k} - y_k\| < \infty$ with probability 1 [cf. (23)]. Therefore, the relation in (26) satisfies the conditions of Theorem 3.1 with $\zeta_k = D_k = 0$, thus implying that $\|w_{j,k} - y_k\|$ converges with probability 1 for every $j \in V$. □

Let us compare Theorem 6.1 and Corollary 5.1. Corollary 5.1 provided sufficient conditions for the different agents to have consensus in the mean. Theorem 6.1 strengthens this to consensus with probability 1 and in mean square sense, for a smaller class of stepsize sequences under a stricter assumption.

We next show that the consensus vector is actually in the optimal set, provided that the optimal set is nonempty and the conditional expectations $\|\mathsf{E}[\epsilon_{i,k+1} \mid F_k]\|$ are diminishing.

Theorem 6.2 *Let Assumptions 2.1, 2.2, 2.3 and 6.1 hold. Suppose that the subgradients of each f_i are uniformly bounded over X, i.e., for each $i \in V$ there is C_i such that*

$$\|\nabla f_i(x)\| \leq C_i \quad \text{for all } x \in X.$$

Also, assume that $\sum_{k=0}^{\infty} \|\mathsf{E}[\epsilon_{i,k+1} \mid F_k]\|^2 < \infty$ for all $i \in V$. Further, let the stepsize sequence $\{\alpha_k\}$ be such that $\sum_{k=1}^{\infty} \alpha_k = \infty$ and $\sum_{k=1}^{\infty} \alpha_k^2 < \infty$. Then, if the optimal

set X^* is nonempty, the iterate sequence $\{w_{i,k}\}$ of each agent $i \in V$ converges to the same optimal point with probability 1 and in mean square.

Proof Observe that the conditions of Lemma 4.2 are satisfied. Letting $z = x^*$ for some $x^* \in X^*$, taking conditional expectations and using the bounds on the error moments, we obtain for any $x^* \in X^*$ and any k, with probability 1,

$$\sum_{i=1}^m \mathsf{E}[\|v_{i,k+1} - x^*\|^2 \mid F_k] \leq \sum_{i=1}^m \|v_{i,k} - x^*\|^2 - 2\alpha_{k+1}(f(y_k) - f^*)$$

$$+ 2\alpha_{k+1}\left(\max_{i \in V} C_i\right) \sum_{j=1}^m \|y_k - w_{j,k}\|$$

$$+ 2\alpha_{k+1} \sum_{i=1}^m \mu_{i,k+1}\|v_{i,k} - x^*\| + \alpha_{k+1}^2 \sum_{i=1}^m (C_i + v_i)^2,$$

where $f^* = f(x^*)$, and we use the notation $\mu_{i,k+1} = \|\mathsf{E}[\epsilon_{i,k+1} \mid F_k]\|$. Using the inequality

$$2\alpha_{k+1}\mu_{i,k+1}\|v_{i,k} - x^*\| \leq \alpha_{k+1}^2\|v_{i,k} - x^*\|^2 + \mu_{i,k+1}^2,$$

we obtain with probability 1,

$$\sum_{i=1}^m \mathsf{E}[\|v_{i,k+1} - x^*\|^2 \mid F_k] \leq \sum_{i=1}^m (1 + \alpha_{k+1}^2)\|v_{i,k} - x^*\|^2$$

$$- 2\alpha_{k+1}\left((f(y_k) - f^*) - \left(\max_{i \in V} C_i\right) \sum_{j=1}^m \|y_k - w_{j,k}\|\right.$$

$$\left. + \sum_{i=1}^m \mu_{i,k+1}^2 - \frac{1}{2}\alpha_{k+1} \sum_{i=1}^m (C_i + v_i)^2\right). \tag{27}$$

By Theorem 6.1, we have with probability 1, $\sum_k \alpha_{k+1}\|w_{j,k} - y_k\| < \infty$. Further, since $\sum_k \mu_{i,k}^2 < \infty$ and $\sum_k \alpha_k^2 < \infty$ with probability 1, the relation in (27) satisfies the conditions of Theorem 3.1. We therefore have

$$\sum_k \alpha_k(f(y_k) - f^*) < \infty, \tag{28}$$

and $\|v_{i,k} - x^*\|$ converges with probability 1 and in mean square. In addition, by Theorem 6.1, we have $\lim_{k \to \infty} \|w_{i,k} - y_k\| = 0$ for all i, with probability 1. Hence, $\lim_{k \to \infty} \|v_{i,k} - y_k\| \to 0$ for all i, with probability 1. Therefore, $\|y_k - x^*\|$ converges with probability 1 for any $x^* \in X^*$. Moreover, from (28) and the fact that $\sum_k \alpha_k = \infty$, by continuity of f, it follows that y_k, and hence $w_{i,k}$, must converge to a vector in X^* with probability 1 and in mean square. $\square$

Theorem 6.2 does not require the compactness of the constraint set X. This is due to the assumption that both the stepsize α_k and the norms $\|\mathsf{E}\left[\epsilon_{i,k+1} \mid F_k\right]\|$ of the conditional errors are square summable. In addition, note that the result of Theorem 6.2 remains valid when the condition $\sum_{k=0}^{\infty} \|\mathsf{E}\left[\epsilon_{i,k+1} \mid F_k\right]\|^2 < \infty$ for all i is replaced with $\sum_{k=0}^{\infty} \alpha_{k+1} \|\mathsf{E}\left[\epsilon_{i,k+1} \mid F_k\right]\| < \infty$ for all i.

7 Discussion

We studied the effects of stochastic subgradient errors on distributed algorithm for network of agents with time-varying connectivity. We first considered very general errors with bounded second moments and obtained explicit bounds on the agent disagreements and on the expected deviation of the limiting function value from the optimal. The bounds are explicitly given as a function of the network properties, objective function and the error moments. For networks that are connected at all times and η is independent of the size of the network, the bound scales as $\alpha(\max_{i \in V}\{C_i + v_i\})^2 m^4$, where m is the number of agents in the network, α is the stepsize limit, and C_i and v_i^2 are respectively the subgradient norm bound and the bound on the second moment of the subgradient errors for agent i. For the constant stepsize case, we obtained a bound on the performance of the algorithm after a finite number of iterations. There, we showed that deviation from the "error-bound" diminishes at rate $\frac{1}{t}$, where t is the number of iterations. Finally, we proved that when the expected error and the stepsize converge to 0 sufficiently fast, the agents reach a consensus and the iterate sequences of agents converge to a common optimal point with probability 1 and in mean square.

We make the following remarks. First, it can be shown that the disagreement results in Corollary 5.1 and Theorem 6.1 hold even when the agents use non-identical stepsizes. However, with non-identical agent stepsizes there is no guarantee that the sum of the objectives rather than a weighted sum, is minimized.

Future work includes several important extensions of the distributed model studied here. At first, we have assumed no communication delays between the agents and synchronous processing. An important extension is to consider the properties of the algorithm in asynchronous networks with communication delays, as in [20]. At second, we assumed perfect communication scenario, i.e., noiseless communication links. In wireless network applications, the links are typically noisy and this has to be taken into consideration. At third, we have considered the class of convex functions. This restricts the number of possible applications for the algorithm. Further research is to develop distributed algorithms when the functions f_i are not convex.

Appendix

Proof of Lemma 3.1 (a) Let $\epsilon > 0$ be arbitrary. Since $\gamma_k \to \gamma$ and for all k, there is an index K such that $|\gamma_k - \gamma| \leq \epsilon$ for all $k \geq K$. For all $k \geq K + 1$, we

have

$$\sum_{\ell=0}^{k} \beta^{k-\ell} \gamma_\ell = \sum_{\ell=0}^{K} \beta^{k-\ell} \gamma_\ell + \sum_{\ell=K+1}^{k} \beta^{k-\ell} \gamma_\ell$$

$$\leq \max_{0 \leq t \leq K} \gamma_t \sum_{\ell=0}^{K} \beta^{k-\ell} + (\gamma + \epsilon) \sum_{\ell=K+1}^{k} \beta^{k-\ell}.$$

Since $\sum_{\ell=K+1}^{k} \beta^{k-\ell} \leq \frac{1}{1-\beta}$ and $\sum_{\ell=0}^{K} \beta^{k-\ell} = \beta^{k-K}(1 + \cdots + \beta^K) \leq \frac{\beta^{k-K}}{1-\beta}$, it follows that

$$\sum_{\ell=0}^{k} \beta^{k-\ell} \gamma_\ell \leq \left(\max_{0 \leq t \leq K} \gamma_t \right) \frac{\beta^{k-K}}{1-\beta} + \frac{\gamma + \epsilon}{1-\beta} \quad \text{for all } k \geq K+1.$$

Therefore, $\limsup_{k \to \infty} \sum_{\ell=0}^{k} \beta^{k-\ell} \gamma_\ell \leq \frac{\gamma+\epsilon}{1-\beta}$. Since ϵ is arbitrary, we have $\limsup_{k \to \infty} \sum_{\ell=0}^{k} \beta^{k-\ell} \gamma_\ell \leq \frac{\gamma}{1-\beta}$. Similarly, we have

$$\sum_{\ell=0}^{k} \beta^{k-\ell} \gamma_\ell \geq \min_{0 \leq t \leq K} \gamma_t \sum_{\ell=0}^{K} \beta^{k-\ell} + (\gamma - \epsilon) \sum_{\ell=K+1}^{k} \beta^{k-\ell}.$$

Thus,

$$\liminf_{k \to \infty} \sum_{\ell=0}^{k} \beta^{k-\ell} \gamma_\ell \geq \liminf_{k \to \infty} \left(\min_{0 \leq t \leq K} \gamma_t \sum_{\ell=0}^{K} \beta^{k-\ell} + (\gamma - \epsilon) \sum_{\ell=K+1}^{k} \beta^{k-\ell} \right).$$

Since $\sum_{\ell=0}^{K} \beta^{k-\ell} \geq \beta^{k-K}$ and $\sum_{\ell=K+1}^{k} \beta^{k-\ell} = \sum_{s=0}^{k-(K+1)} \beta^s$, which tends to $1/(1-\beta)$ as $k \to \infty$, it follows that

$$\liminf_{k \to \infty} \sum_{\ell=0}^{k} \beta^{k-\ell} \gamma_\ell \geq \left(\min_{0 \leq t \leq K} \gamma_t \right) \lim_{k \to \infty} \beta^{k-K} + (\gamma - \epsilon) \lim_{k \to \infty} \sum_{s=0}^{k-(K+1)} \beta^s = \frac{\gamma - \epsilon}{1-\beta}.$$

Since ϵ is arbitrary, we have $\liminf_{k \to \infty} \sum_{\ell=0}^{k} \beta^{k-\ell} \gamma_\ell \geq \frac{\gamma}{1-\beta}$. This and $\limsup_{k \to \infty} \sum_{\ell=0}^{k} \beta^{k-\ell} \gamma_\ell \leq \frac{\gamma}{1-\beta}$ (established earlier), imply $\lim_{k \to \infty} \sum_{\ell=0}^{k} \beta^{k-\ell} \gamma_\ell = \frac{\gamma}{1-\beta}$.

(b) Let $\sum_{k=0}^{\infty} \gamma_k < \infty$. For any integer $M \geq 1$, we have

$$\sum_{k=0}^{M} \left(\sum_{\ell=0}^{k} \beta^{k-\ell} \gamma_\ell \right) = \sum_{\ell=0}^{M} \gamma_\ell \sum_{t=0}^{M-\ell} \beta^t \leq \sum_{\ell=0}^{M} \gamma_\ell \frac{1}{1-\beta},$$

implying that

$$\sum_{k=0}^{\infty} \left(\sum_{\ell=0}^{k} \beta^{k-\ell} \gamma_\ell \right) \leq \frac{1}{1-\beta} \sum_{\ell=0}^{\infty} \gamma_\ell < \infty.$$

(c) Since $\limsup_{k\to\infty}\gamma_k=\gamma$, for every $\epsilon>0$ there is a large enough K such that $\gamma_k\le\gamma+\epsilon$ for all $k\ge K$. Thus, for any $M>K$,

$$\frac{\sum_{k=0}^{M}\gamma_k\zeta_k}{\sum_{k=0}^{M}\zeta_k}=\frac{\sum_{k=0}^{K}\gamma_k\zeta_k}{\sum_{k=0}^{M}\zeta_k}+\frac{\sum_{k=K+1}^{M}\gamma_k\zeta_k}{\sum_{k=0}^{M}\zeta_k}\le\frac{\sum_{k=0}^{K}\gamma_k\zeta_k}{\sum_{k=0}^{M}\zeta_k}+(\gamma+\epsilon)\frac{\sum_{k=K+1}^{M}\zeta_k}{\sum_{k=0}^{M}\zeta_k}.$$

By letting $M\to\infty$ and using $\sum_k\zeta_k=\infty$, we see that $\limsup_{M\to\infty}\frac{\sum_{k=0}^{M}\gamma_k\zeta_k}{\sum_{k=0}^{M}\zeta_k}\le\gamma+\epsilon$, and since ϵ is arbitrary, the result for the limit superior follows.

Analogously, if $\liminf_{k\to\infty}\gamma_k=\gamma$, then for every $\epsilon>0$ there is a large enough K such that $\gamma_k\ge\gamma-\epsilon$ for all $k\ge K$. Thus, for any $M>K$,

$$\frac{\sum_{k=0}^{M}\gamma_k\zeta_k}{\sum_{k=0}^{M}\zeta_k}=\frac{\sum_{k=0}^{K}\gamma_k\zeta_k}{\sum_{k=0}^{M}\zeta_k}+\frac{\sum_{k=K+1}^{M}\gamma_k\zeta_k}{\sum_{k=0}^{M}\zeta_k}\ge\frac{\sum_{k=0}^{K}\gamma_k\zeta_k}{\sum_{k=0}^{M}\zeta_k}+(\gamma-\epsilon)\frac{\sum_{k=K+1}^{M}\zeta_k}{\sum_{k=0}^{M}\zeta_k}.$$

Letting $M\to\infty$ and using $\sum_k\zeta_k=\infty$, we obtain $\liminf_{M\to\infty}\frac{\sum_{k=0}^{M}\gamma_k\zeta_k}{\sum_{k=0}^{M}\zeta_k}\ge\gamma-\epsilon$. Since $\epsilon>0$ is arbitrary, we have $\liminf_{M\to\infty}\frac{\sum_{k=0}^{M}\gamma_k\zeta_k}{\sum_{k=0}^{M}\zeta_k}\ge\gamma$. This relation and the relation for the limit superior yield $\lim_{M\to\infty}\frac{\sum_{k=0}^{M}\gamma_k\zeta_k}{\sum_{k=0}^{M}\zeta_k}=\gamma$ when $\gamma_k\to\gamma$. $\square$

References

1. Nedić, A., Bertsekas, D.P.: Incremental subgradient method for nondifferentiable optimization. SIAM J. Optim. **12**, 109–138 (2001)
2. Nedić, A., Bertsekas, D.P.: Convergence rate of incremental algorithms. In: Uryasev, S., Pardalos, P.M. (eds.) Stochastic Optimization: Algorithms and Applications, pp. 223–264. Kluwer Academic, Dordrecht (2001)
3. Pardalos, P.M., Resende, M.G.C.: Handbook of Applied Optimization. Oxford University Press, New York (2002)
4. Nedić, A., Ozdaglar, A.: Distributed subgradient methods for multi-agent optimization. Trans. Automat. Control **54**, 48–61 (2009)
5. Nedić, A., Ozdaglar, A.: On the rate of convergence of distributed asynchronous subgradient methods for multi-agent optimization. In: Proceedings of the 46th IEEE Conference on Decision and Control, pp. 4711–4716 (2007)
6. Sundhar, R.S., Nedić, A., Veeravalli, V.V.: Incremental stochastic subgradient algorithms for convex optimization. SIAM J. Optim. **20**, 691–717 (2009)
7. Nedić, A., Ozdaglar, A., Parrilo, P.A.: Constrained consensus and optimization in multi-agent networks. Lab. for Information and Decision Systems Technical Report 2779, Massachusetts Institute of Technology (2008)
8. Rabbat, M.G., Nowak, R.D.: Distributed optimization in sensor networks. In: Proceedings of International Symposium on Information Processing in Sensor Networks, pp. 20–27 (2004)
9. Blatt, D., Hero, A.O., Gauchman, H.: A convergent incremental gradient method with constant stepsize. SIAM J. Optim. **18**, 29–51 (2007)
10. Johansson, B., Rabi, M., Johansson, M.: A simple peer-to-peer algorithm for distributed optimization in sensor networks. In: Proceedings of the 46th IEEE Conference on Decision and Control, pp. 4705–4710 (2007)
11. Johansson, B.: On distributed optimization in networked systems. Ph.D. thesis, Royal Institute of Technology, Stockholm (2008)
12. Sundhar, R.S.: Distributed optimization in multi-agent systems with applications to distributed regression. Ph.D. thesis, University of Illinois at Urbana-Champaign (2009)

13. Nedić, A., Ozdaglar, A.: Cooperative distributed multi-agent optimization. In: Daniel, P.P., Eldar, Y.C. (eds.) Convex Optimization in Signal Processing and Communications, pp. 340–386. Cambridge University Press, Cambridge (2010)

14. Rabbat, M.G., Nowak, R.D.: Quantized incremental algorithms for distributed optimization. IEEE J. Sel. Areas Commun. **23**, 798–808 (2005)

15. Gaivoronski, A.A.: Convergence properties of backpropagation for neural nets via theory of stochastic gradient methods. Optim. Methods Softw. **4**, 117–134 (1994)

16. Solodov, M.V., Zavriev, S.K.: Error stability properties of generalized gradient-type algorithms. J. Optim. Theory Appl. **98**, 663–680 (1998)

17. Bertsekas, D.P., Tsitsiklis, J.N.: Gradient convergence in gradient methods with errors. SIAM J. Optim. **10**, 627–642 (2000)

18. Kiwiel, K.C.: Convergence of approximate and incremental subgradient methods for convex optimization. SIAM J. Optim. **14**, 807–840 (2003)

19. Nedić, A., Bertsekas, D.P.: The effect of deterministic noise in sub-gradient methods. Math. Program. (2009). doi:10.1007/s10107-008-0262-5

20. Tsitsiklis, J.N.: Problems in decentralized decision making and computation. Ph.D. thesis, Massachusetts Institute of Technology (1984)

21. Tsitsiklis, J.N., Bertsekas, D.P., Athans, M.: Distributed asynchronous deterministic and stochastic gradient optimization algorithms. IEEE Trans. Automat. Control **31**, 803–812 (1986)

22. Bertsekas, D.P., Tsitsiklis, J.N.: Parallel and Distributed Computation: Numerical Methods. Athena Scientific, Belmont (1997)

23. Jadbabaie, A., Lin, J., Morse, S.: Coordination of groups of mobile autonomous agents using nearest neighbor rules. IEEE Trans. Automat. Control **48**, 998–1001 (2003)

24. Spanos, D.S., Olfati-Saber, R., Murray, R.M.: Approximate distributed Kalman filtering in sensor networks with quantifiable performance. In: Proceedings of IEEE International Conference on Information Processing in Sensor Networks, pp. 133–139 (2005)

25. Kar, S., Moura, J.M.F.: Distributed consensus algorithms in sensor networks: link and channel noise. IEEE Trans. Signal Process. **57**, 355–369 (2009)

26. Xiao, L., Boyd, S., Kim, S.-J.: Distributed average consensus with least mean square deviation. J. Parallel Distrib. Comput. **67**, 33–46 (2007)

27. Lobel, I., Ozdaglar, A.: Distributed subgradient methods over random networks. Lab. for Information and Decision Systems Technical Report 2800, Massachusetts Institute of Technology (2008)

28. Olshevsky, A., Tsitsiklis, J.N.: Convergence speed in distributed consensus and averaging. SIAM J. Control Optim. **48**, 33–55 (2009)

29. Nedić, A., Olshevsky, A., Ozdaglar, A., Tsitsiklis, J.N.: Distributed subgradient algorithms and quantization effects. In: Proceedings of the 47th IEEE Conference on Decision and Control, pp. 4177–4184 (2008)

30. Huang, M., Manton, J.H.: Stochastic approximation for consensus seeking: mean square and almost sure convergence. In: Proceedings of the 46th IEEE Conference on Decision and Control, pp. 306–311 (2007)

31. Ermoliev, Y.: Stochastic Programming Methods. Nauka, Moscow (1976)

32. Ermoliev, Y.: Stochastic quasi-gradient methods and their application to system optimization. Stochastics **9**, 1–36 (1983)

33. Ermoliev, Y.: Stochastic quazigradient methods. In: Ermoliev, Y., Wets, R.J.-B. (eds.) Numerical Techniques for Stochastic Optimization, pp. 141–186. Springer, New York (1988)

34. Sundhar, R.S., Veeravalli, V.V., Nedić, A.: Distributed and non-autonomous power control through distributed convex optimization. In: Proceedings of the 28th IEEE Conference on Computer Communications INFOCOM, pp. 3001–3005 (2009)

35. Bertsekas, D.P., Nedić, A., Ozdaglar, A.: Convex Analysis and Optimization. Athena Scientific, Belmont (2003)

36. Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton (1970)

37. Polyak, B.T.: Introduction to Optimization. Optimization Software, New York (1987)

38. Billingsley, P.: Probability and Measure. Wiley, New York (1979)