Asynchronous Gossip Algorithm for Stochastic Optimization: Constant Stepsize Analysis*

S. Sundhar Ram[1], Angelia Nedić[2], and Venu V. Veeravalli[3]

[1] Electrical and Computer Engineering Department, University of Illinois at Urbana-Champaign, Urbana IL 61801, USA ssriniv5@illinois.edu
[2] Industrial and Enterprise Systems Engineering Department, University of Illinois at Urbana-Champaign, Urbana IL 61801, USA angelia@illinois.edu
[3] Electrical and Computer Engineering Department, University of Illinois at Urbana-Champaign, Urbana IL 61801, USA vvv@illinois.edu

Summary. We consider the problem of minimizing the sum of convex functions over a network when each component function is known (with stochastic errors) to a specific network agent. We discuss a gossip based algorithm of [2], and we analyze its error bounds for a constant stepsize that is uncoordinated across the agents.

1 Introduction

The gossip optimization algorithm proposed in [2] minimizes a sum of functions when each component function is known (with stochastic errors) to a specific network agent. The algorithm is reliant on the gossip-consensus scheme of [1], which serves as a main mechanism for the decentralization of the overall network optimization problem. The gossip-based optimization algorithm is *distributed* and *totally asynchronous* since there is no central coordinator and the agents do not have a common notion of time. Furthermore, the algorithm is *completely local* since each agent knows only its neighbors, and relies on its own local information and some limited information received from its neighbors. Agents have no information about the global network.

In [2], the convergence properties of the algorithm with a (random) diminishing uncoordinated stepsize was studied. In this paper we study the properties of the algorithm when the agents use deterministic *uncoordinated constant stepsizes*. Our primary interest is in establishing the limiting error bounds for the method. We provide such error bounds for strongly convex functions and for general convex functions (through the use of the running averages of the iterates). The bounds are given explicitly in terms of the problem data, the network connectivity parameters and the agent stepsize values. The bounds scale linearly in the number of agents.

* This work is supported by NSF Career grant 0742538.

2 Problem, algorithm and assumptions

Throughout this paper, we use $\|x\|$ to denote the Euclidean norm of a vector x. We write $\mathbf{1}$ to denote the vector with all entries equal to 1. The matrix norm $\|M\|$ of a matrix M is the norm induced by the Euclidean vector norm. We use x^T and M^T to denote the transpose of a vector x and a matrix M, respectively. We write $[x]_i$ to denote the i-th component of a vector x. Similarly, we write $[M]_{i,j}$ or $M_{i,j}$ to indicate the (i,j)-th component of a matrix M. We write $|S|$ to denote the cardinality of a set S with finitely many elements.

Consider a network of m agents that are indexed by $1,\ldots,m$, and let $V = \{1,\ldots,m\}$. The agents communicate over a network with a static topology represented by an undirected graph $(V,\mathscr{E})$, where $\mathscr{E}$ is the set of undirected links $\{i,j\}$ with $i \neq j$ and $\{i,j\} \in \mathscr{E}$ only if agents i and j can communicate.

We are interested in solving the following problem over the network:

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

where each f_i is a function defined over the set $X \subseteq \mathbb{R}^n$. The problem (1) is to be solved under the following restrictions on the network information. Each agent i knows only its own objective function f_i and it can compute the (sub)gradients ∇f_i with stochastic errors. Furthermore, each agent can communicate and exchange some information with its local neighbors only.

To solve problem (1), we consider an algorithm that is based on the gossip consensus model in [1]. Let $N(i)$ be the set of all neighbors of agent i, i.e., $N(i) = \{j \in V \mid \{i,j\} \in \mathscr{E}\}$. Each agent has its local clock that ticks at a Poisson rate of 1 independently of the clocks of the other agents. At each tick of its clock, agent i communicates with a randomly selected neighbor $j \in N(i)$ with probability $P_{ij} > 0$, where $P_{ij} = 0$ for $j \notin N(i)$. Then, agent i and the selected neighbor j exchange their current estimates of the optimal solution, and each of these agents performs an update using the received estimate and the erroneous (sub)gradient direction of its objective function.

Consider a single virtual clock that ticks whenever any of the local Poisson clocks tick. Let Z_k be the time of the k-th tick of the virtual Poisson clock, and let the time be discretized according to the intervals $[Z_{k-1}, Z_k)$, $k \geq 1$. Let I_k denote the index of the agent that wakes up at time k, and let J_k denote the index of a neighbor that is selected for communication. Let $x_{i,k}$ denote the iterate of agent i at time k. The iterates are generated according to the following rule. Agents other than I_k and J_k do not update:

$$x_{i,k} = x_{i,k-1} \qquad \text{for } i \notin \{I_k, J_k\}. \tag{2}$$

Agents I_k and J_k average their current iterate and update independently using subgradient steps as follows:

$$v_{i,k} = \left(x_{I_k,k-1} + x_{J_k,k-1}\right)/2,$$
$$x_{i,k} = P_X[v_{i,k} - \alpha_i(\nabla f_i(v_{i,k}) + \epsilon_{i,k})], \quad (3)$$

where P_X denotes the Euclidean projection on the set X, $\nabla f_i(x)$ is a subgradient of f_i at x, α_i is a positive stepsize, and $\epsilon_{i,k}$ is *stochastic error* in computing $\nabla f_i(v_{i,k})$. The updates are initialized with random vectors $x_{i,0}$, $i \in V$, which are assumed to be mutually independent and also independent of all the other random variables in the process.

The key difference between the work in [2] and this paper is in the stepsize. The work in [2] considers a diminishing (random) stepsize $\alpha_{i,k}$, which is defined in terms of the frequency of agent i updates. In contrast, in this paper, we consider the method with a deterministic constant stepsize $\alpha_i > 0$ for all i. As the stepsizes across agents need not be the same, the algorithm does not require any coordination among the agents.

We next discuss our assumptions.

Assumption 1 *The underlying communication graph $(V, \mathscr{E})$ is connected.*

Assumption 1 ensures that, through the gossip strategy, the information of each agent reaches every other agent frequently enough. However, to ensure that the common vector solves problem (1), some additional assumptions are needed for the set X and the functions f_i. We use the following.

Assumption 2 *The set $X \subseteq \mathbb{R}^n$ is compact and convex. Each function f_i is defined and convex over an open set containing the set X.*

Differentiability of the functions f_i is not assumed. At points where the gradient does not exist, we use a subgradient. Under the compactness of X, the subgradients are uniformly bounded over X, i.e., for some $C > 0$ we have

$$\sup_{x \in X} \|\nabla f_i(x)\| \leq C \qquad \text{for all } i \in V.$$

Furthermore, the following *approximate subgradient relation* holds:

$$\nabla f_i(v)^T(v - x) \geq f_i(y) - f_i(x) - C\|v - y\| \quad \text{for any } x, y, v \in X \text{ and } i \in V. \quad (4)$$

We now discuss the random errors $\epsilon_{i,k}$ in computing the subgradients $\nabla f_i(x)^T$ at points $x = v_{i,k}$. Let $\mathcal{F}_k$ be the σ-algebra generated by the entire history of the algorithm up to time k inclusively, i.e.,

$$\mathcal{F}_k = \{x_{i,0},\ i \in V\} \cup \{I_\ell, J_\ell, \epsilon_{I_\ell,\ell}, \epsilon_{J_\ell,\ell};\ 1 \leq \ell \leq k\} \qquad \text{for all } k \geq 1,$$

where $\mathcal{F}_0 = \{x_{i,0},\ i \in V\}$. We use the following assumption on the errors.

Assumption 3 *With probability 1, for all $i \in \{I_k, J_k\}$ and $k \geq 1$, the errors satisfy $\mathsf{E}\left[\epsilon_{i,k} \mid \mathcal{F}_{k-1}, I_k, J_k\right] = 0$ and $\mathsf{E}\left[\|\epsilon_{i,k}\|^2 \mid \mathcal{F}_{k-1}, I_k, J_k\right] \leq \nu^2$ for some ν.*

When X and each f_i are convex, every vector $v_{i,k}$ is a convex combination of $x_{j,k} \in X$ (see Eq. (3)), implying that $v_{i,k} \in X$. In view of subgradient boundedness and Assumption 3, it follows that for $k \geq 1$,

$$\mathsf{E}\left[\|\nabla f_i^T(v_{i,k}) + \epsilon_{i,k}\|^2 \mid \mathcal{F}_{k-1}, I_k, J_k\right] \leq (C + \nu)^2 \quad \text{for } i \in \{I_k, J_k\}. \quad (5)$$

3 Preliminaries

We provide an alternative description of the algorithm, and study the properties of the agent's disagreements. Define the matrix W_k as follows:

$$W_k = I - \frac{1}{2}(e_{I_k} - e_{J_k})(e_{I_k} - e_{J_k})^T \qquad \text{for all } k, \tag{6}$$

where $e_i \in \mathbb{R}^m$ has its i-th entry equal to 1, and the other entries equal to 0. Using W_k, we can write method (2)–(3) as follows: for all $k \geq 1$ and $i \in V$,

$$\begin{aligned} x_{i,k} &= v_{i,k} + p_{i,k}\chi_{\{i\in\{I_k,J_k\}\}}, \\ v_{i,k} &= \sum_{j=1}^{m}[W_k]_{ij}\; x_{j,k-1}, \\ p_{i,k} &= P_X[v_{i,k} - \alpha_i\left(\nabla f_i(v_{i,k}) + \epsilon_{i,k}\right)] - v_{i,k}, \end{aligned} \tag{7}$$

where $\chi_{\mathscr{C}}$ is the characteristic function of an event $\mathscr{C}$. The matrices W_k are symmetric and stochastic, implying that each $\mathsf{E}\left[W_k\right]$ is doubly stochastic. Thus, by the definition of the method in (7), we can see that

$$\sum_{i=1}^{m}\mathsf{E}\left[\|v_{i,k} - x\|^2 \mid \mathcal{F}_{k-1}\right] \leq \sum_{j=1}^{m}\|x_{j,k-1} - x\|^2 \qquad \text{for all } x \in \mathbb{R}^n \text{ and } k, \tag{8}$$

$$\sum_{i=1}^{m}\mathsf{E}\left[\|v_{i,k} - x\| \mid \mathcal{F}_{k-1}\right] \leq \sum_{j=1}^{m}\|x_{j,k-1} - x\| \qquad \text{for all } x \in \mathbb{R}^n \text{ and } k. \tag{9}$$

In our analysis, we use the fact that $W_k^2 = W_k$, $(W_k - \frac{1}{m}\mathbf{1}\mathbf{1}^T)^2 = W_k - \frac{1}{m}\mathbf{1}\mathbf{1}^T$ and that the norm of the matrices $\mathsf{E}[W_k] - \frac{1}{m}\mathbf{1}\mathbf{1}^T$ is equal to the second largest eigenvalue of $\mathsf{E}[W_k]$. We let λ denote the square of this eigenvalue, i.e., $\lambda = \|\mathsf{E}[W_k] - \frac{1}{m}\mathbf{1}\mathbf{1}^T\|^2$. We have the following lemma.

Lemma 1. *Let Assumption 1 hold. Then, we have $\lambda < 1$.*

We next provide an estimate for the disagreement among the agents.

Lemma 2. *Let Assumptions 1–3 hold*[4]*, and let $\{x_{i,k}\}$, $i = 1, \ldots, m$, be the iterate sequences generated by algorithm (7). Then, we have for all i,*

$$\limsup_{k\to\infty}\sum_{i=1}^{m}\mathsf{E}\left[\|x_{i,k} - \bar{y}_k\|\right] \leq \frac{\sqrt{2m}\,\bar{\alpha}}{1-\sqrt{\lambda}}\,(C + \nu),$$

where $\bar{y}_k = \frac{1}{m}\sum_{j=1}^{m} x_{j,k}$ for all k, and $\bar{\alpha} = \max_{1\leq j\leq m}\, \alpha_j$.

[4] Here, we only need the error boundedness from Assumption 3.

Proof. We will consider coordinate-wise relations by defining the vector $z_k^\ell \in \mathbb{R}^m$, for each $\ell \in \{1, \ldots, n\}$, as the vector with entries $[x_{i,k}]_\ell$, $i = 1, \ldots, m$. From the definition of the method in (7), we have

$$z_k^\ell = W_k\, z_{k-1}^\ell + \zeta_k^\ell \qquad \text{for } k \geq 1, \tag{10}$$

where $\zeta_k^\ell \in \mathbb{R}^m$ is a vector with coordinates $[\zeta_k^\ell]_i$ given by

$$[\zeta_k^\ell]_i = \begin{cases} [P_X[v_{i,k} - \alpha_i\,(\nabla f_i(v_{i,k}) + \epsilon_{i,k})] - v_{i,k}]_\ell & \text{if } i \in \{I_k, J_k\}, \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

Furthermore, note that $[\bar{y}_k]_\ell$ is the average of the entries of the vector z_k^ℓ, i.e.,

$$[\bar{y}_k]_\ell = \frac{1}{m}\,\mathbf{1}^T z_k^\ell \qquad \text{for all } k \geq 0. \tag{12}$$

By Eqs. (10) and (12), we have $[\bar{y}_k]_\ell = \frac{1}{m}\left(\mathbf{1}^T W_k z_{k-1}^\ell + \mathbf{1}^T \zeta_k^\ell\right)$, implying

$$\begin{aligned} z_k^\ell - \mathbf{1}[\bar{y}_k]_\ell &= W_k z_{k-1}^\ell + \zeta_k^\ell - \frac{1}{m}\,\mathbf{1}\mathbf{1}^T(W_k z_{k-1}^\ell + \zeta_k^\ell) \\ &= \left(W_k - \frac{1}{m}\,\mathbf{1}\mathbf{1}^T\right) z_{k-1}^\ell + \left(I - \frac{1}{m}\,\mathbf{1}\mathbf{1}^T\right)\zeta_k^\ell, \end{aligned}$$

where I denotes the identity matrix, and the last equality follow by the doubly stochasticity of W_k, i.e., $\mathbf{1}^T W_k = \mathbf{1}^T$. Since the matrices W_k are stochastic, i.e., $W_k \mathbf{1} = \mathbf{1}$, it follows $\left(W_k - \frac{1}{m}\,\mathbf{1}\mathbf{1}^T\right)\mathbf{1} = 0$, implying that $\left(W_k - \frac{1}{m}\,\mathbf{1}\mathbf{1}^T\right)[\bar{y}_{k-1}]_\ell \mathbf{1} = 0$. Hence,

$$z_k^\ell - [\bar{y}_k]_\ell \mathbf{1} = D_k(z_{k-1}^\ell - [\bar{y}_{k-1}]_\ell \mathbf{1}) + M\zeta_k^\ell \qquad \text{for all } k \geq 1,$$

where $D_k = W_k - \frac{1}{m}\,\mathbf{1}\mathbf{1}^T$ and $M = I - \frac{1}{m}\,\mathbf{1}\mathbf{1}^T$. Thus, we have for $\ell = 1, \ldots, n$ and all $k \geq 1$,

$$\|z_k^\ell - [\bar{y}_k]_\ell \mathbf{1}\|^2 \leq \|D_k(z_{k-1}^\ell - [\bar{y}_{k-1}]_\ell \mathbf{1})\|^2 + \|M\zeta_k^\ell\|^2 + 2\|D_k(z_{k-1}^\ell - [\bar{y}_{k-1}]_\ell \mathbf{1})\|\,\|M\zeta_k^\ell\|.$$

By summing these relations over $\ell = 1, \ldots, n$, and then taking the expectation and using Hölder's inequality we obtain for all $k \geq 1$,

$$\sum_{\ell=1}^n \mathsf{E}\left[\|z_k^\ell - [\bar{y}_k]_\ell \mathbf{1}\|^2\right] \leq \left(\sqrt{\sum_{\ell=1}^n \mathsf{E}\left[\|D_k(z_{k-1}^\ell - [\bar{y}_{k-1}]_\ell \mathbf{1})\|^2\right]} + \sqrt{\sum_{\ell=1}^n \mathsf{E}\left[\|M\zeta_k^\ell\|^2\right]}\right)^2 \tag{13}$$

Using the fact the matrix W_k is independent of the past $\mathcal{F}_{k-1}$, we have

$$\sum_{\ell=1}^n \mathsf{E}\left[\left\|D_k(z_{k-1}^\ell - [\bar{y}_{k-1}]_\ell \mathbf{1})\right\|^2 \mid \mathcal{F}_{k-1}\right] \leq \lambda \sum_{\ell=1}^n \|z_{k-1}^\ell - [\bar{y}_{k-1}]_\ell \mathbf{1}\|^2, \tag{14}$$

where $\lambda = \|\mathsf{E}[D_k^T D_k]\|^2 = \|\mathsf{E}[D_k]\|^2$, and $\lambda < 1$ from Lemma 1.

We next estimate the second term in (13). The matrix $M = I - \frac{1}{m}\mathbf{1}\mathbf{1}^T$ is a projection matrix (it projects on the subspace orthogonal to the vector $\mathbf{1}$), so that we have $\|M\|^2 = 1$, implying that $\|M\zeta_k^\ell\|^2 \le \|\zeta_k^\ell\|^2$ for all k. Using this and the definition of ζ_k^ℓ in (11), we obtain

$$\|M\zeta_k^\ell\|^2 \le 2 \sum_{i\in\{I_k,J_k\}} \left|[P_X[v_{i,k} - \alpha_i\left(\nabla f_i(v_{i,k}) + \epsilon_{i,k}\right)] - v_{i,k}]_\ell\right|^2.$$

Therefore,

$$\sum_{\ell=1}^n \mathsf{E}\left[\|M\zeta_k^\ell\|^2\right] \le 2\mathsf{E}\left[\mathsf{E}\left[\sum_{i\in\{I_k,J_k\}} \alpha_i^2\|\nabla f_i(v_{i,k}) + \epsilon_{i,k}\|^2 \mid \mathcal{F}_{k-1}, I_k, J_k\right]\right]$$
$$\le 2\bar{\alpha}^2(C+\nu)^2,$$

where in the last inequality we use $\bar{\alpha} = \max_i \alpha_i$ and relation (5). Combining the preceding relation with Eqs. (13) and (14), we obtain

$$\sqrt{\sum_{\ell=1}^n \mathsf{E}\left[\|z_k^\ell - [\bar{y}_k]_\ell\mathbf{1}\|^2\right]} \le \sqrt{\lambda}\sqrt{\sum_{\ell=1}^n \mathsf{E}\left[\|z_{k-1}^\ell - [\bar{y}_{k-1}]_\ell\mathbf{1}\|^2\right]} + \sqrt{2}\,\bar{\alpha}(C+\nu).$$

Since $\lambda < 1$, by recursively using the preceding relation, we have

$$\limsup_{k\to\infty} \sqrt{\sum_{\ell=1}^n \mathsf{E}\left[\|z_k^\ell - [\bar{y}_k]_\ell\mathbf{1}\|^2\right]} \le \frac{\sqrt{2}\,\bar{\alpha}}{1-\sqrt{\lambda}}(C+\nu).$$

The result now follows by $\sum_{i=1}^m \mathsf{E}\left[\|x_{i,k} - \bar{y}_k\|^2\right] = \sum_{\ell=1}^n \mathsf{E}\left[\|z_k^\ell - \mathbf{1}[\bar{y}_k]_\ell\|^2\right]$ and $\sum_{i=1}^m \mathsf{E}\left[\|x_{i,k} - \bar{y}_k\|\right] \le \sqrt{m}\,\sqrt{\sum_{i=1}^m \mathsf{E}\left[\|x_{i,k} - \bar{y}_k\|^2\right]}$. $\square$

The bound in Lemma 2 captures the dependence of the differences between $x_{i,k}$ and their current average $\bar{y}_k$ in terms of the maximum stepsize and the communication graph. The impact of the communication graph $(V,\mathscr{E})$ is captured by the spectral radius λ of the expected matrices $\mathsf{E}\left[(W_k - \frac{1}{m}\mathbf{1}\mathbf{1}^T)^2\right]$.

4 Error Bounds

We have the following result for strongly convex functions.

Proposition 1. *Let Assumptions 1–3 hold. Let each function f_i be strongly convex over the set X with a constant σ_i, and let α_i be such that $2\alpha_i\sigma_i < 1$. Then, for the sequences $\{x_{i,k}\}$, $i \in V$, generated by (7), we have for all i,*

$$\limsup_{k\to\infty} \sum_{i=1}^m \mathsf{E}[\|x_k^i - x^*\|^2] \le \frac{\bar{\omega} - \underline{\omega}}{1-q}\, 2mCC_X + \frac{\bar{\alpha}\bar{\omega}}{1-q}\left(m + \frac{2\sqrt{2m}}{1-\sqrt{\lambda}}\right)(C+\nu)^2,$$

where x^* is the optimal solution of problem (1), $q = \max_i\{1 - 2\gamma_i\alpha_i\sigma_i\}$, $\gamma_i = \frac{1}{m}\left(1 + \sum_{j\in N(i)} P_{ji}\right)$, $C_X = \max_{x,y\in X}\|x-y\|$, $\bar{\alpha} = \max_i \alpha_i$, $\bar{\omega} = \max_i \gamma_i\alpha_i$, and $\underline{\omega} = \min_i \gamma_i\alpha_i$.

Proof. The sum $f = \sum_{i=1}^m f_i$ is strongly convex with constant $\sigma = \sum_{i=1}^m \sigma_i$. Thus, problem (1) has a unique optimal solution $x^* \in X$. From relation (7), the nonexpansive property of the projection operation, and relation (5) we obtain for the optimal point x^*, and any k and $i \in \{I_k, J_k\}$,

$$\mathsf{E}\left[\|x_{i,k} - x^*\|^2 \mid \mathcal{F}_{k-1}, I_k, J_k\right] \le \|v_{i,k} - x^*\|^2 - 2\alpha_i \nabla f_i(v_{i,k})^T (v_{i,k} - x^*) \\ + \alpha_i^2 (C+\nu)^2. \tag{15}$$

By the strong convexity of f_i, it follows

$$\nabla f_i(v_{i,k})^T(v_{i,k} - x^*) \ge \sigma_i \|v_{i,k} - x^*\|^2 + \nabla f_i(x^*)^T (v_{i,k} - x^*).$$

Using $\bar{y}_{k-1} = \frac{1}{m}\sum_{j=1}^m x_{j,k-1}$, we have $\nabla f_i(x^*)^T(v_{i,k} - x^*) = \nabla f_i(x^*)^T(\bar{y}_{k-1} - x^*) + \nabla f_i(x^*)^T(v_{i,k} - \bar{y}_{k-1})$, which in view of $\|\nabla f_i(x^*)\| \le C$ implies

$$\nabla f_i(x^*)^T(v_{i,k} - x^*) \ge \nabla f_i(x^*)^T(\bar{y}_{k-1} - x^*) - C\,\|v_{i,k} - \bar{y}_{k-1}\|. \tag{16}$$

By combining the preceding two relations with inequality (15), we obtain

$$\mathsf{E}\left[\|x_{i,k} - x^*\|^2 \mid \mathcal{F}_{k-1}, I_k, J_k\right] \le (1 - 2\alpha_i\sigma_i)\|v_{i,k} - x^*\|^2 \\ + \alpha_i^2(C+\nu)^2 - 2\alpha_i \nabla f_i(x^*)^T(\bar{y}_{k-1} - x^*) + 2\alpha_i C \|v_{i,k} - \bar{y}_{k-1}\|.$$

Taking the expectation with respect to F_{k-1} and using the fact the preceding inequality holds with probability γ_i (the probability that agent i updates at time k), and $x_{i,k} = v_{i,k}$ with probability $1 - \gamma_i$, we obtain for any i and k,

$$\mathsf{E}\left[\|x_{i,k} - x^*\|^2 \mid \mathcal{F}_{k-1}\right] \le (1 - 2\gamma_i\alpha_i\sigma_i)\mathsf{E}\left[\|v_{i,k} - x^*\|^2 \mid \mathcal{F}_{k-1}\right] \\ + \gamma_i\alpha_i^2(C+\nu)^2 - 2\gamma_i\alpha_i\nabla f_i(x^*)^T(\bar{y}_{k-1} - x^*) + 2\gamma_i\alpha_i C\,\mathsf{E}\left[\|v_{i,k} - \bar{y}_{k-1}\| \mid \mathcal{F}_{k-1}\right].$$

Adding and subtracting $(\min_i \gamma_i\alpha_i)\,\nabla f_i(x^*)^T(\bar{y}_{k-1} - x^*)$, and using $\bar{y}_{k-1} \in X$ and the compactness of X, we obtain

$$\mathsf{E}\left[\|x_{i,k} - x^*\|^2 \mid \mathcal{F}_{k-1}\right] \le q\mathsf{E}\left[\|v_{i,k} - x^*\|^2 \mid \mathcal{F}_{k-1}\right] + \bar{\omega}\bar{\alpha}(C+\nu)^2 \\ + 2(\bar{\omega} - \underline{\omega})CC_X - 2\underline{\omega}\nabla f_i(x^*)^T(\bar{y}_{k-1} - x^*) + 2\bar{\omega}C\mathsf{E}\left[\|v_{i,k} - \bar{y}_{k-1}\| \mid \mathcal{F}_{k-1}\right],$$

where $q = \max_i\{1 - 2\gamma_i\alpha_i\sigma_i\}$, $\underline{\omega} = \min_i \gamma_i\alpha_i$, $\bar{\omega} = \max_i \gamma_i\alpha_i$, $\bar{\alpha} = \max_i \alpha_i$ and $C_X = \max_{x,y\in X}\|x - y\|$. Now, by summing the preceding relations over i, by using $\sum_{i=1}^m \nabla f_i(x^*)^T(\bar{y}_{k-1} - x^*) \ge 0$ and using relation (8) (with $x = x^*$) and relation (9) (with $x = \bar{y}_{k-1}$), we obtain

$$\sum_{i=1}^m \mathsf{E}[\|x_{i,k} - x^*\|^2] \le q\sum_{j=1}^m \mathsf{E}[\|x_{j,k} - x^*\|^2] + m\bar{\omega}\bar{\alpha}(C+\nu)^2 \\ + 2m(\bar{\omega} - \underline{\omega})CC_X + 2\bar{\omega}C\sum_{j=1}^m \mathsf{E}[\|x_{j,k} - \bar{y}_{k-1}\|].$$

The desired estimate follows from the preceding relation by noting that $q < 1$, by taking the limit superior and by using Lemma 2 and $C(C+\nu) \le (C+\nu)^2$. □

Proposition 1 requires each node to select a stepsize α_i so that $2\alpha_i\sigma_i < 1$, which can be done since each node knows its strong convexity constant σ_i. Furthermore, note that the relation $q = \max_{1\le i\le m}\{1-\gamma_i\alpha_i\sigma_i\} < 1$ can be ensured globally over the network without any coordination among the agents.

The following error estimate holds without strong convexity.

Proposition 2. *Let Assumptions 1–3 hold. Then, for the sequences* $\{x_{i,k}\}$, $i \in V$, *generated by (7), we have for all* i,

$$\limsup_{k\to\infty} \frac{1}{k}\sum_{t=1}^{k} \mathsf{E}[f(x_{i,t-1})] \le f^* + m\,(\rho-1)\,CC_X + \bar{\alpha}\left((\rho+m)\frac{\sqrt{2m}}{1-\sqrt{\lambda}} + \frac{m}{2}\rho\right)(C+\nu)^2,$$

where f^* *is the optimal value of problem (1),* $C_X = \max_{x,y\in X}\|x-y\|$, $\rho = \frac{\max_i \gamma_i\alpha_i}{\min_i \gamma_i\alpha_i}$, $\gamma_i = \frac{1}{m}\left(1+\sum_{j\in N(i)} P_{ji}\right)$ *and* $\bar{\alpha} = \max_i \alpha_i$.

Proof. The optimal set X^* is nonempty. Thus, Eq. (15) holds for any $x^* \in X^*$. From approximate subgradient relation (4) it follows

$$\nabla f_i(v_{i,k})^T(v_{i,k}-x^*) \ge f_i(\bar{y}_{k-1}) - f_i(x^*) - C\|v_{i,k}-\bar{y}_{k-1}\|.$$

The preceding relation and Eq. (15) yield for all $i \in \{I_k, J_k\}$ and $k \ge 1$,

$$\mathsf{E}\left[\|x_{i,k}-x^*\|^2 \mid \mathcal{F}_{k-1}, I_k, J_k\right] \le \|v_{i,k}-x^*\|^2 - 2\alpha_i(f_i(\bar{y}_{k-1})-f_i(x^*)) + 2\alpha_i C\|v_{i,k}-\bar{y}_{k-1}\| + \alpha_i^2(C+\nu)^2,$$

where $C_X = \max_{x,y\in X}\|x-y\|$. The preceding relation holds when $i \in \{I_k, J_k\}$, which happens with probability γ_i. When $i \notin \{I_k, J_k\}$, we have $x_{i,k} = v_{i,k}$ (see Eq. (7)), which happens with probability $1-\gamma_i$. Thus, by taking the expectation conditioned on $\mathcal{F}_{k-1}$, we obtain

$$\mathsf{E}\left[\|x_{i,k}-x^*\|^2 \mid \mathcal{F}_{k-1}\right] \le \mathsf{E}\left[\|v_{i,k}-x^*\|^2 \mid \mathcal{F}_{k-1}\right] - 2\gamma_i\alpha_i(f_i(\bar{y}_{k-1})-f_i(x^*)) + 2\gamma_i\alpha_i C\mathsf{E}\left[\|v_{i,k}-\bar{y}_{k-1}\| \mid \mathcal{F}_{k-1}\right] + \gamma_i\alpha_i^2(C+\nu)^2.$$

Letting $\underline{\omega} = \min_{1\le i\le m}\{\gamma_i\alpha_i\}$ and $\bar{\omega} = \max_{1\le i\le m}\{\gamma_i\alpha_i\}$, and using

$$|f_i(\bar{y}_{k-1}) - f_i(x^*)| \le C\|\bar{y}_{k-1}-x^*\| \le CC_X,$$

which holds by the subgradient boundedness and the fact $\bar{y}_k \in X$, we see that

$$\mathsf{E}\left[\|x_{i,k}-x^*\|^2 \mid \mathcal{F}_{k-1}\right] \le \mathsf{E}\left[\|v_{i,k}-x^*\|^2 \mid \mathcal{F}_{k-1}\right] - 2\underline{\omega}(f_i(\bar{y}_{k-1})-f_i(x^*)) + 2(\bar{\omega}-\underline{\omega})CC_X + 2\bar{\omega}C\mathsf{E}\left[\|v_{i,k}-\bar{y}_{k-1}\| \mid \mathcal{F}_{k-1}\right] + \bar{\omega}\bar{\alpha}(C+\nu)^2,$$

where $\bar{\alpha} = \max_{1\le i\le m}\alpha_i$. By summing the preceding inequalities over i, and using Eq. (8) with $x = x^*$ and Eq. (9) with $x = \bar{y}_{k-1} \in X$, we obtain

$$2\underline{\omega}\mathsf{E}[f(\bar{y}_{k-1}) - f(x^*)] \le \sum_{j=1}^{m} \mathsf{E}[\|x_{j,k-1} - x^*\|^2] - \sum_{i=1}^{m} \mathsf{E}[\|x_{i,k} - x^*\|^2]$$
$$+2m(\bar{\omega} - \underline{\omega})CC_X + 2\bar{\omega}C\sum_{j=1}^{m} \mathsf{E}[\|x_{j,k-1} - \bar{y}_{k-1}\|] + m\bar{\omega}\bar{\alpha}(C+\nu)^2,$$

where $f = \sum_{i=1}^{m} f_i$. Next, after dividing the preceding relation by $2\underline{\omega}$ and noting that by convexity and the boundedness of the subgradients of each f_i, we have

$$f(x_{i,k-1}) - f^* \le f(\bar{y}_{k-1}) - f^* + mC\|x_{i,k-1} - \bar{y}_{k-1}\|,$$

we obtain for all i,

$$\mathsf{E}[f(x_{i,k-1}) - f(x^*)] \le \frac{1}{2\underline{\omega}}\left(\sum_{j=1}^{m} \mathsf{E}[\|x_{j,k-1} - x^*\|^2] - \sum_{i=1}^{m} \mathsf{E}[\|x_{i,k} - x^*\|^2]\right)$$
$$+m(\rho-1)CC_X + (\rho+m)\,C\sum_{j=1}^{m} \mathsf{E}[\|x_{j,k-1} - \bar{y}_{k-1}\|] + \frac{m}{2}\rho\bar{\alpha}(C+\nu)^2,$$

where $\rho = \frac{\bar{\omega}}{\underline{\omega}}$. By summing these relations from time 1 to time k, and then averaging with respect to k, we obtain

$$\frac{1}{k}\sum_{t=1}^{k} \mathsf{E}[f(x_{i,t-1}) - f(x^*)] \le \frac{1}{2k\underline{\omega}}\left(\sum_{j=1}^{m} \mathsf{E}[\|x_{j,0} - x^*\|^2] - \sum_{i=1}^{m} \mathsf{E}[\|x_{i,k} - x^*\|^2]\right)$$
$$+m(\rho-1)CC_X + (\rho+m)C\frac{1}{k}\sum_{t=1}^{k}\sum_{j=1}^{m} \mathsf{E}[\|x_{j,t-1} - \bar{y}_{t-1}\|] + \frac{m}{2}\rho\bar{\alpha}(C+\nu)^2.$$

Letting $k \to \infty$ and using the relation

$$\limsup_{k\to\infty} \frac{1}{k}\sum_{t=1}^{k}\left(\sum_{j=1}^{m} \mathsf{E}[\|x_{j,k-1} - \bar{y}_{k-1}\|]\right) \le \limsup_{k\to\infty} \sum_{j=1}^{m} \mathsf{E}[\|x_{j,k-1} - \bar{y}_{k-1}\|],$$

we have for any i,

$$\limsup_{k\to\infty} \frac{1}{k}\sum_{t=1}^{k} \mathsf{E}[f(x_{i,t-1}) - f(x^*)] \le m(\rho-1)CC_X$$
$$+(\rho+m)C\limsup_{k\to\infty}\sum_{j=1}^{m} \mathsf{E}[\|x_{j,k-1} - \bar{y}_{k-1}\|] + \frac{m}{2}\rho\bar{\alpha}(C+\nu)^2.$$

By Lemma 2 we have

$$\limsup_{k\to\infty}\sum_{j=1}^{m} \mathsf{E}[\|x_{j,k-1} - \bar{y}_{k-1}\|] \le \frac{\sqrt{2m}\,\bar{\alpha}}{1-\sqrt{\lambda}}\,(C+\nu),$$

which together with the preceding relation yields for all i,

$$\limsup_{k\to\infty} \frac{1}{k}\sum_{t=1}^{k} \left(\mathsf{E}[f(x_{i,t-1})] - f(x^*)\right) \le m\,(\rho-1)\,CC_X$$
$$+(\rho+m)C\frac{\sqrt{2m}\,\bar{\alpha}}{1-\sqrt{\lambda}}\,(C+\nu) + \frac{m}{2}\rho\bar{\alpha}(C+\nu)^2.$$

By using $C(C+\nu) \le (C+\nu)^2$ and grouping the terms accordingly, we obtain the desired relation. □

By Proposition 2 and the convexity of f, we have for $u_{i,k} = \frac{1}{k}\sum_{t=1}^{k} x_{i,t-1}$,

$$\limsup_{k\to\infty} \mathsf{E}[f\left(u_{i,k}\right)] \le f^* + B,$$

where $B = m\,(\rho-1)\,CC_X + \bar{\alpha}\left((\rho+m)\frac{\sqrt{2m}}{1-\sqrt{\lambda}} + \frac{m}{2}\rho\right)(C+\nu)^2$. When the ratio $\rho = \frac{\max_i \gamma_i\alpha_i}{\min_i \gamma_i\alpha_i}$ is close to value 1, the bound is approximately given by $B \approx \bar{\alpha}\left((1+m)\frac{\sqrt{2m}}{1-\sqrt{\lambda}} + \frac{m}{2}\right)(C+\nu)^2$. In this case, the bound scales in the size m of the network as $m^{3/2}$, which is by order $1/2$ less than the scaling of the bound for the distributed consensus-based subgradient algorithm of [3], which scales at best as m^2.

5 Discussion

The bounds scale well with the size of the network. For strongly convex functions, the bound in Proposition 1 scales independently of the size of the network if the degrees of the nodes are about the same order and do not change with the size of the network. The bound in Proposition 2 scales as $m\sqrt{m}$ with the size m of the network. In our development, we have assumed that the network topology is static, which may not be realistic in some applications. Of future interest is to investigate the algorithm for dynamic network topology.

References

1. Boyd S, Ghosh A, Prabhakar B, Shah D (2006) Randomized gossip algorithms. IEEE Trans. on Information Theory 52:2508–2530
2. Sundhar RS, Nedić A, Veeravalli VV (2009) Asynchronous Gossip Algorithm for stochastic optimization. Proc. of IEEE Conf. on Decision and Control
3. Sundhar RS, Veeravalli VV, Nedić A (2009) Distributed and non-autonomous power control through distributed convex optimization. IEEE INFOCOM