Online Appendices of "Distributed Multi-Agent Optimization with State-Dependent Communication"

Ilan Lobel, Asuman Ozdaglar, Diego Feijer

B Appendix - Omitted Proofs

Proof of Proposition 1 Consider a network consisting of two agents solving a one-dimensional minimization problem. The first agent's objective function is $f_1(x) = -x$, while the second agent's objective function is $f_2(x) = 2x$. Both agents' feasible sets are equal to $X_1 = X_2 = [0, \infty)$. Let $x_1(0) \geq x_2(0) \geq 0$. The elements of the communication matrix are given by

$$a_{1,2}(k) = a_{2,1}(k) = \begin{cases} \gamma, \text{ with probability} & \min\left\{\delta, \frac{1}{|x_1(k)-x_2(k)|^C}\right\}; \\ 0, \text{ with probability} & 1 - \min\left\{\delta, \frac{1}{|x_1(k)-x_2(k)|^C}\right\}, \end{cases}$$

for some $\gamma \in (0, 1/2]$ and $\delta \in [1/2, 1)$.

The optimal solution set of this multi-agent optimization problem is the singleton $X^* = \{0\}$ and the optimal solution is $f^* = 0$. We now prove that $\lim_{k\to\infty} x_1(k) = \infty$ with probability 1 implying that $\lim_{k\to\infty} |f(x_1(k)) - f^*| = \infty$.

From the iteration in Eq. (2), we have that for any k,

$$x_1(k+1) = a_{1,1}(k)x_1(k) + a_{1,2}(k)x_2(k) + \alpha \quad (37)$$

$$x_2(k+1) = \max\{0, a_{2,1}(k)x_1(k) + a_{2,2}(k)x_2(k) - 2\alpha\}. \quad (38)$$

We do not need to project $x_1(k+1)$ onto $X_1 = [0,\infty)$ because $x_1(k+1)$ is non-negative if $x_1(k)$ and $x_2(k)$ are both non-negative. Note that since $\gamma \leq 1/2$, this iteration preserves $x_1(k) \geq x_2(k) \geq 0$ for all $k \in \mathbb{N}$.

We now show that for any $k \in \mathbb{N}$ and any $x_1(k) \geq x_2(k) \geq 0$, there is probability at least $\epsilon > 0$ that the two agents will never communicate again, i.e.,

$$P(a_{1,2}(k') = a_{2,1}(k') = 0 \text{ for all } k' \geq k | x(k)) \geq \epsilon > 0. \quad (39)$$

If the agents do not communicate on periods $k, k+1, ..., k+j-1$ for some $j \geq 1$, then

$$\begin{aligned} x_1(k+j) - x_2(k+j) &= (x_1(k+j) - x_1(k)) + (x_1(k) - x_2(k)) + (x_2(k) - x_2(k+j)) \\ &\geq \alpha j + 0 + 0, \end{aligned}$$

from Eqs. (37) and (38) and the fact that $x_1(k) \geq x_2(k)$. Therefore, the communication probability at period $k+j$ can be bounded by

$$P(a_{1,2}(k+j) = 0 | x(k), a_{1,2}(k') = 0 \text{ for all } k' \in \{k, ..., k+j-1\}) \geq 1 - \min\{\delta, (\alpha j)^{-C}\}.$$

Applying this bound recursively for all $j \geq k$, we obtain

$$\begin{aligned} &P(a_{1,2}(k') = 0 \text{ for all } k' \geq k | x(k)) \\ &= \prod_{j=0}^{\infty} P(a_{1,2}(k+j) = 0 | x(k), a_{1,2}(k') = 0 \text{ for all } k' \in \{k, ..., k+j-1\}) \\ &\geq \prod_{j=0}^{\infty} \left(1 - \min\{\delta, (\alpha j)^{-C}\}\right) \end{aligned}$$

for all k and all $x_1(k) \geq x_2(k)$. We now show that $\prod_{j=0}^{\infty} \left(1 - \min\{\delta, (\alpha j)^{-C}\}\right) > 0$ if $C > 1$. Define the constant $\overline{K} = \left\lceil \frac{2^{\frac{1}{C}}}{\alpha} \right\rceil$. Since $\delta \geq 1/2$, we have that $(\alpha j)^{-C} \leq \delta$ for $j \geq \overline{K}$. Hence, we can separate the infinite product into two components:

$$\prod_{j=0}^{\infty} \left(1 - \min\{\delta, (\alpha j)^{-C}\}\right) \geq \left[\prod_{j<\overline{K}} \left(1 - \min\{\delta, (\alpha j)^{-C}\}\right)\right] \left[\prod_{j \geq \overline{K}} \left(1 - (\alpha j)^{-C}\right)\right].$$

Note that the term in the first brackets in the equation above is a product of a finite number of strictly positive numbers and, therefore, is a strictly positive number. We, thus, have to show only that $\prod_{j\geq\overline{K}}\left(1-(\alpha j)^{-C}\right)>0$. We can bound this product by

$$\prod_{j\geq\overline{K}}\left(1-(\alpha j)^{-C}\right)=\exp\left(\log\left(\prod_{j\geq\overline{K}}\left(1-(\alpha j)^{-C}\right)\right)\right)$$
$$=\exp\left(\sum_{j\geq\overline{K}}\log\left(1-(\alpha j)^{-C}\right)\right)\geq\exp\left(\sum_{j\geq\overline{K}}-(\alpha j)^{-C}\log(4)\right),$$

where the inequality follows from $\log(x)\geq(x-1)\log(4)$ for all $x\in[1/2,1]$. Since $C>1$, the sum $\sum_{j\geq\overline{K}}(\alpha j)^{-C}$ is finite and $\prod_{j=0}^{\infty}\left(1-\min\{\delta,(\alpha j)^{-C}\}\right)>0$, yielding Eq. (39).

Let K^* be the (random) set of periods when agents communicate, i.e., $a_{1,2}(k)=a_{2,1}(k)=\gamma$ if and only if $k\in K^*$. For any value $k\in K^*$ and any $x_1(k)\geq x_2(k)$, there is probability at least ϵ that the agents do not communicate after k. Conditionally on the state, this is an event independent of the history of the algorithm by the Markov property. If K^* has infinitely many elements, then by the Borel-Cantelli Lemma we obtain that, with probability 1, for infinitely many k's in K^* there is no more communication between the agents after period k. This contradicts the infinite cardinality of K^*. Hence, the two agents only communicate finitely many times and $\lim_{k\to\infty}x_1(k)=\infty$ with probability 1. □

Proof of Lemma 1 Letting $s=0$ in Eq. (7) yields,

$$x_i(k)=\sum_{j=1}^{m}[\Phi(k-1,0)]_{ij}x_j(0)$$
$$-\sum_{r=1}^{k-1}\sum_{j=1}^{m}[\Phi(k-1,r)]_{ij}\alpha(r-1)d_j(r-1)-\alpha(k-1)d_i(k-1)$$
$$+\sum_{r=1}^{k-1}\sum_{j=1}^{m}[\Phi(k-1,r)]_{ij}e_j(r-1)+e_i(k-1).$$

Since the matrices $A(k)$ are doubly stochastic with probability one for all k (cf. Assumption 3), it follows that the transition matrices $\Phi(k,s)$ are doubly stochastic for all $k\geq s\geq 0$, implying that every entry $[\Phi(k,s)]_{ij}$ belongs to $[0,1]$ with probability one. Thus, for all k we have,

$$\|x_i(k)\|\leq\sum_{j=1}^{m}\|x_j(0)\|+\sum_{r=1}^{k-1}\sum_{j=1}^{m}\alpha(r-1)\|d_j(r-1)\|+\alpha(k-1)\|d_i(k-1)\|$$
$$+\sum_{r=1}^{k-1}\sum_{j=1}^{m}\|e_j(r-1)\|+\|e_i(k-1)\|.$$

Using the bound L on the subgradients, this implies

$$\|x_i(k)\|\leq\sum_{j=1}^{m}\|x_j(0)\|+\sum_{r=0}^{k-1}mL\alpha(r)+\sum_{r=0}^{k-1}\sum_{j=1}^{m}\|e_j(r)\|.$$

Finally, the fact that $\|x_i(k)-x_h(k)\|\leq\|x_i(k)\|+\|x_h(k)\|$ for every $i,h\in\mathcal{M}$, establishes the desired result. □

Proof of Lemma 2 (a) The proof is based on the fact that the communication matrices $A(k)$ are Markovian on the state $x(k)$, for all time $k\geq 0$. First, note that

$$P(G(k)|x(s)=\overline{x})=P\left(\bigcap_{l=1}^{m-1}\left(B_l(k)\cap D_l(k)\right)\middle|x(s)=\overline{x}\right)$$
$$=P\left(\bigcap_{l=1}^{m-1}B_l(k)\middle|x(s)=\overline{x}\right)P\left(\bigcap_{l=1}^{m-1}D_l(k)\middle|\bigcap_{l=1}^{m-1}B_l(k),x(s)=\overline{x}\right).\quad(40)$$

To simplify notation, let $W = 2m(L+M)$. We show that for all $k \geq s$,

$$\inf_{\overline{x} \in R_M(s)} P\left(\bigcap_{l=1}^{m-1} B_l(k) \middle| x(s) = \overline{x}\right) \geq \min\left\{\delta, \frac{K}{(\Delta + W \sum_{r=1}^{k+2m-3} \alpha(r))^C}\right\}^{(m-1)}. \quad (41)$$

We skip the proof of the equivalent bound for the second term in Eq. (40) to avoid repetition. By conditioning on $x(k)$ we obtain for all $k \geq s$,

$$\inf_{\overline{x} \in R_M(s)} P\left(\bigcap_{l=1}^{m-1} B_l(k) \middle| x(s) = \overline{x}\right) =$$
$$\inf_{\overline{x} \in R_M(s)} \int_{x' \in \mathbb{R}^{m \times n}} P\left(\bigcap_{l=1}^{m-1} B_l(k) \middle| x(k) = x', x(s) = \overline{x}\right) dP(x(k) = x' | x(s) = \overline{x}).$$

Using the Markov Property, we see that conditional on $x(s)$ can be removed from the right-hand side probability above, since $x(k)$ already contains all relevant information with respect to $\cap_{l=1}^{m-1} B_l(k)$. By the definition of $R_M(\cdot)$ [see Eq. (9)], if $x(s) \in R_M(s)$, then $x(k) \in R_M(k)$ for all $k \geq s$ with probability 1. Therefore,

$$\inf_{\overline{x} \in R_M(s)} P\left(\bigcap_{l=1}^{m-1} B_l(k) \middle| x(s) = \overline{x}\right) \geq \inf_{\overline{x} \in R_M(k)} P\left(\bigcap_{l=1}^{m-1} B_l(k) \middle| x(k) = x'\right). \quad (42)$$

By the definition of $B_1(k)$,

$$\inf_{\overline{x} \in R_M(k)} P\left(\bigcap_{l=1}^{m-1} B_l(k) \middle| x(k) = \overline{x}\right) = \quad (43)$$
$$\inf_{\overline{x} \in R_M(k)} P(a_{e_1}(k) \geq \gamma | x(s) = \overline{x}) P\left(\bigcap_{l=2}^{m-1} B_l(k) \middle| a_{e_1}(k) \geq \gamma, x(k) = \overline{x}\right).$$

Define

$$Q(k) = \min\left\{\delta, \frac{K}{\left(\Delta + W \sum_{r=1}^{k} \alpha(r)\right)^C}\right\},$$

and note that, in view of the assumption imposed on the norm of the projection errors and based on Lemma 1, we get

$$\max_{i,h \in \mathcal{M}} \|x_i(k) - x_h(k)\| \leq \Delta + W \sum_{r=0}^{k-1} \alpha(r).$$

Hence, from Eq. (6) we have

$$P(a_{ij}(k) \geq \gamma | x(k) = \overline{x}) \geq Q(k). \quad (44)$$

Thus, combining Eqs. (43) and (44) we obtain,

$$\inf_{\overline{x} \in R_M(k)} P\left(\bigcap_{l=1}^{m-1} B_l(k) \middle| x(k) = \overline{x}\right) \geq Q(k) \inf_{\overline{x} \in R_M(k)} P\left(\bigcap_{l=2}^{m-1} B_l(k) \middle| a_{e_1}(k) \geq \gamma, x(k) = \overline{x}\right). \quad (45)$$

By conditioning on the state $x(k+1)$, and repeating the use of the Markov property and the definition of $R_M(k+1)$, we can bound the right-hand side of the equation above,

$$\inf_{\overline{x} \in R_M(k)} P\left(\bigcap_{l=2}^{m-1} B_l(k) \middle| a_{e_1}(k) \geq \gamma, x(k) = \overline{x}\right)$$
$$= \inf_{\overline{x} \in R_M(k)} \int_{x'} P\left(\bigcap_{l=2}^{m-1} B_l(k) \middle| x(k+1) = x'\right) dP(x(k+1) = x' | a_{e_1}(k) \geq \gamma, x(k) = \overline{x})$$
$$\geq \inf_{x' \in R_M(k+1)} P\left(\bigcap_{l=2}^{m-1} B_l(k) \middle| x(k+1) = x'\right). \quad (46)$$

Combining Eqs. (43), (45) and (46), we obtain

$$\inf_{\overline{x}\in R_M(k)} P\left(\bigcap_{l=1}^{m-1} B_l(k)\middle| x(k)=\overline{x}\right) \geq Q(k) \inf_{\overline{x}\in R_M(k+1)} P\left(\bigcap_{l=2}^{m-1} B_l(k)\middle| x(k+1)=x'\right).$$

Repeating this process for all $l=1,...,m-1$, this yields

$$\inf_{\overline{x}\in R_M(k)} P\left(\bigcap_{l=1}^{m-1} B_l(k)\middle| x(k)=\overline{x}\right) \geq \prod_{l=1}^{m-1} Q(k+l-1).$$

Since Q is a decreasing function, $\prod_{l=1}^{m-1} Q(k+l-1) \geq Q(k+2m-3)^{m-1}$. Combining with Eq. (42), we have that for all $k \geq s$

$$\inf_{\overline{x}\in R_M(s)} P\left(\bigcap_{l=1}^{m-1} B_l(k)\middle| x(s)=\overline{x}\right) \geq Q(k+2m-3)^{m-1},$$

producing the desired Eq. (41).

(b) Let $G^c(k)$ represent the complement of $G(k)$. Note that

$$P\left(\bigcup_{l=0}^{u-1} G(k+2(m-1)l)\middle| x(k)=\overline{x}\right) = 1 - P\left(\bigcap_{l=0}^{u-1} G^c(k+2(m-1)l)\middle| x(k)=\overline{x}\right).$$

By conditioning on $G^c(k)$, we obtain

$$P\left(\bigcap_{l=0}^{u-1} G^c(k+2(m-1)l)\middle| x(k)=\overline{x}\right) =$$
$$P\left(G^c(k)|x(k)=\overline{x}\right) P\left(\bigcap_{l=1}^{u-1} G^c(k+2(m-1)l)\middle| G^c(k), x(k)=\overline{x}\right).$$

We bound the term $P\left(G^c(k)|x(k)=\overline{x}\right)$ using the result from part (a). We bound the second term in the right-hand side of the equation above using the Markov property and the definition of $R_M(\cdot)$, which is the same technique from part (a),

$$\sup_{\overline{x}\in R_M(k)} P\left(\bigcap_{l=1}^{u-1} G^c(k+2(m-1)l)\middle| G^c(k), x(k)=\overline{x}\right)$$
$$= \sup_{\overline{x}\in R_M(k)} \int_{x'} P\left(\bigcap_{l=1}^{u-1} G^c(k+2(m-1)l)\middle| x(k+2(m-1))=x'\right) \times$$
$$dP(x(k+2(m-1))=x'|G^c(k), x(k)=\overline{x})$$
$$\leq \sup_{\overline{x}\in R_M(k+2(m-1))} P\left(\bigcap_{l=1}^{u-1} G^c(k+2(m-1)l)\middle| x(k+2(m-1))=x'\right).$$

The result follows by repeating the bound above u times. ◻

Proof of Lemma 4 From Assumption 5, with $p=0$, we obtain that there exists some $\overline{K}\in\mathbb{N}$ such that $\alpha(k)\leq 1/k$ for all $k\geq\overline{K}$. Therefore,

$$\sum_{k=0}^{\infty}\alpha^2(k) \leq \sum_{k=0}^{\overline{K}-1}\alpha^2(k) + \sum_{k=\overline{K}}^{\infty}\frac{1}{k^2} \leq \overline{K}\max_{k\in\{0,\dots,\overline{K}-1\}}\alpha^2(k) + \frac{\pi^2}{6} < \infty.$$

Hence, $\{\alpha(k)\}_{k\in\mathbb{N}}$ is square summable. Now, let $\overline{\alpha}(k) = \frac{1}{(k+2)\log(k+2)}$ for all $k\in\mathbb{N}$. This sequence of stepsizes satisfies Assumption 5 and is not summable since for all $K'\in\mathbb{N}$

$$\sum_{k=0}^{K'}\overline{\alpha}(k) \geq \log(\log(K'+2))$$

and $\lim_{K'\to\infty} \log(\log(K'+2)) = \infty$. □

Proof of Proposition 3 From Assumption 2, we have that there exists a set of edges $\mathcal{E}$ of the strongly connected graph $(\mathcal{M}, \mathcal{E})$ such that for all $(j,i) \in \mathcal{E}$, all $k \geq 0$ and all $\overline{x} \in \mathbb{R}^{m\times n}$,

$$P(a_{ij}(k) \geq \gamma | x(k) = \overline{x}) \geq \min\left\{\delta, \frac{K}{\|\overline{x}_i - \overline{x}_j\|^C}\right\}.$$

The function $\min\left\{\delta, \frac{K}{\|\overline{x}_i - \overline{x}_j\|^C}\right\}$ is continuous and, therefore, it attains its optimum when minimized over the compact set $\prod_{i\in\mathcal{M}} X_i$, i.e.,

$$\inf_{\overline{x}\in\prod_{i\in\mathcal{M}} X_i} \min\left\{\delta, \frac{K}{\|\overline{x}_i - \overline{x}_j\|^C}\right\} = \min_{\overline{x}\in\prod_{i\in\mathcal{M}} X_i} \min\left\{\delta, \frac{K}{\|\overline{x}_i - \overline{x}_j\|^C}\right\}.$$

Since the function $\min\left\{\delta, \frac{K}{\|\overline{x}_i - \overline{x}_j\|^C}\right\}$ is strictly positive for any $\overline{x} \in \mathbb{R}^{m\times n}$, we obtain that there exists some positive ϵ such that

$$\epsilon = \inf_{\overline{x}\in\prod_{i\in\mathcal{M}} X_i} \min\left\{\delta, \frac{K}{\|\overline{x}_i - \overline{x}_j\|^C}\right\} > 0.$$

Hence, for all $(j,i) \in \mathcal{E}$, all $k \geq 0$ and all $\overline{x} \in \prod_{i\in\mathcal{M}} X_i$,

$$P(a_{ij}(k) \geq \gamma | x(k) = \overline{x}) \geq \epsilon. \tag{47}$$

Since there is a uniform bound on the probability of communication for any given edge in $\mathcal{E}$ that is independent of the state $x(k)$, we can use an extended version of Lemma 7 from [14]. In particular, Lemma 7 as stated in [14] requires the communication probability along edges to be independent of $x(k)$ which does not apply here, however, it can be extended with straightforward modifications to hold if the independence assumption were to be replaced by the condition specified in Eq. (47), implying the desired result. □

Proof of Lemma 5 (a) Since $x_i(k+1) = P_{X_i}[v_i(k) - \alpha(k)d_i(k)]$, it follows from the property of the projection error $e_i(k)$ in Eq. (36) that for any $z \in X$,

$$\|x_i(k+1) - z\|^2 \leq \|v_i(k) - \alpha(k)d_i(k) - z\|^2 - \|e_i(k)\|^2.$$

By expanding the term $\|v_i(k) - \alpha(k)d_i(k) - z\|^2$, we obtain

$$\|v_i(k) - \alpha(k)d_i(k) - z\|^2 = \|v_i(k) - z\|^2 + \alpha^2(k)\|d_i(k)\|^2 - 2\alpha(k)d_i(k)'(v_i(k) - z).$$

Since $v_i(k) = \sum_{j=1}^m a_{ij}(k)x_j(k)$, using the convexity of the norm square function and the stochasticity of the weights $a_{ij}(k)$, $j = 1, \ldots, m$, it follows that

$$\|v_i(k) - z\|^2 \leq \sum_{j=1}^m a_{ij}(k)\|x_j(k) - z\|^2.$$

Combining the preceding relations, we obtain

$$\begin{aligned}\|x_i(k+1) - z\|^2 \leq & \sum_{j=1}^m a_{ij}(k)\|x_j(k) - z\|^2 + \alpha^2(k)\|d_i(k)\|^2 \\ & -2\alpha(k)d_i(k)'(v_i(k) - z) - \|e_i(k)\|^2.\end{aligned}$$

By summing the preceding relation over $i = 1, \ldots, m$, and using the doubly stochasticity of the weights, i.e.,

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

we obtain the desired result.

(b) Since $d_i(k)$ is a subgradient of $f_i(x)$ at $x = v_i(k)$, we have

$$d_i(k)'(v_i(k) - z) \geq f_i(v_i(k)) - f_i(z).$$

Combining this with the inequality in part (a), using subgradient boundedness and dropping the nonpositive projection error term on the right handside, we obtain

$$\sum_{i=1}^{m} \|x_i(k+1) - z\|^2 \leq \sum_{i=1}^{m} \|x_i(k) - z\|^2 + \alpha^2(k) m L^2 - 2\alpha(k) \sum_{i=1}^{m} (f_i(v_i(k)) - f_i(z)),$$

proving the first claim. This relation implies that

$$\begin{aligned}\sum_{j=1}^{m} \|x_j(k+1) - z\|^2 \leq \sum_{j=1}^{m} \|x_j(k) - z\|^2 &+ \alpha^2(k) m L^2 - 2\alpha(k) \sum_{i=1}^{m} \left(f_i(v_i(k)) - f_i(y(k))\right) \\ &- 2\alpha(k) \left(f(y(k)) - f(z)\right).\end{aligned} \tag{48}$$

In view of the subgradient boundedness and the stochasticity of the weights, it follows

$$|f_i(v_i(k)) - f_i(y(k))| \leq L\|v_i(k) - y(k)\| \leq L \sum_{j=1}^{m} a_{ij}(k) \|x_j(k) - y(k)\|,$$

implying, by the doubly stochasticity of the weights, that

$$\sum_{i=1}^{m} |f_i(v_i(k)) - f_i(y(k))| \leq L \sum_{j=1}^{m} \left(\sum_{i=1}^{m} a_{ij}(k) \right) \|x_j(k) - y(k)\| = L \sum_{j=1}^{m} \|x_j(k) - y(k)\|.$$

By using this in relation (48), we see that for any $z \in X$, and all i and k,

$$\begin{aligned}\sum_{j=1}^{m} \|x_j(k+1) - z\|^2 \leq \sum_{j=1}^{m} \|x_j(k) - z\|^2 &+ \alpha^2(k) m L^2 + 2\alpha(k) L \sum_{j=1}^{m} \|x_j(k) - y(k)\| \\ &- 2\alpha(k) \left(f(y(k)) - f(z)\right).\end{aligned}$$

□

Proof of Lemma 6 From Eq. (7), we have for all i and $k \geq s$,

$$\begin{aligned} x_i(k+1) = \sum_{j=1}^{m} [\Phi(k,s)]_{ij} x_j(s) &- \sum_{r=s}^{k-1} \sum_{j=1}^{m} [\Phi(k, r+1)]_{ij} \alpha(r) d_j(r) - \alpha(k) d_i(k) \\ &+ \sum_{r=s}^{k-1} \sum_{j=1}^{m} [\Phi(k, r+1)]_{ij} e_j(r) + e_i(k). \end{aligned}$$

Similarly, using relation (28), we can write for $y(k+1)$ and for all k and s with $k \geq s$,

$$y(k+1) = y(s) - \frac{1}{m} \sum_{r=s}^{k-1} \sum_{j=1}^{m} \alpha(r) d_j(r) - \frac{\alpha(k)}{m} \sum_{i=1}^{m} d_i(k) + \frac{1}{m} \sum_{r=s}^{k-1} \sum_{j=1}^{m} e_j(r) + \frac{1}{m} \sum_{j=1}^{m} e_j(k).$$

Therefore, since $y(s) = \frac{1}{m} \sum_{j=1}^{m} x_j(s)$, we have for $s = 0$,

$$\begin{aligned} \|x_i(k) - y(k)\| \leq & \sum_{j=1}^{m} \left| [\Phi(k-1, 0)]_{ij} - \frac{1}{m} \right| \|x_j(0)\| \\ &+ \sum_{r=0}^{k-2} \sum_{j=1}^{m} \left| [\Phi(k-1, r+1)]_{ij} - \frac{1}{m} \right| \alpha(r) \|d_j(r)\| \\ &+ \alpha(k-1) \|d_i(k-1)\| + \frac{\alpha(k-1)}{m} \sum_{j=1}^{m} \|d_j(k-1)\| \end{aligned}$$

$$+\sum_{r=0}^{k-2}\sum_{j=1}^{m}\left|[\Phi(k-1,r+1)]_{ij}-\frac{1}{m}\right|\|e_j(r)\|$$
$$+\|e_i(k-1)\|+\frac{1}{m}\sum_{j=1}^{m}\|e_j(k-1)\|.$$

Using the metric $\rho(k,s)=\max_{i,j\in\mathcal{M}}\left|[\Phi(k,s)]_{ij}-\frac{1}{m}\right|$ for $k\geq s\geq 0$ [cf. Eq. (8)], and the subgradient boundedness, we obtain for all i and $k\geq 2$,

$$\|x_i(k)-y(k)\|\leq m\rho(k-1,0)\sum_{j=1}^{m}\|x_j(0)\|+mL\sum_{r=0}^{k-2}\rho(k-1,r+1)\alpha(r)+2\alpha(k-1)L$$
$$+\sum_{r=0}^{k-2}\rho(k-1,r+1)\sum_{j=1}^{m}\|e_j(r)\|+\|e_i(k-1)\|+\frac{1}{m}\sum_{j=1}^{m}\|e_j(k-1)\|,$$

completing the proof. □

Proof of Lemma 7 Let $\epsilon>0$ be arbitrary. Since $\gamma_k\to 0$, there is an index K such that $\gamma_k\leq\epsilon$ for all $k\geq K$. For all $k\geq K+1$, we have

$$\sum_{\ell=0}^{k}\beta_{k-\ell}\gamma_\ell=\sum_{\ell=0}^{K}\beta_{k-\ell}\gamma_\ell+\sum_{\ell=K+1}^{k}\beta_{k-\ell}\gamma_\ell\leq\max_{0\leq t\leq K}\gamma_t\sum_{\ell=0}^{K}\beta_{k-\ell}+\epsilon\sum_{\ell=K+1}^{k}\beta_{k-\ell}.$$

Since $\sum_{l=0}^{\infty}\beta_l<\infty$, there exists $B>0$ such that $\sum_{\ell=K+1}^{k}\beta_{k-\ell}=\sum_{\ell=0}^{k-K-1}\beta_\ell\leq B$ for all $k\geq K+1$. Moreover, since $\sum_{\ell=0}^{K}\beta_{k-\ell}=\sum_{\ell=k-K}^{k}\beta_\ell$, it follows that for all $k\geq K+1$,

$$\sum_{\ell=0}^{k}\beta_{k-\ell}\gamma_\ell\leq\max_{0\leq t\leq K}\gamma_t\sum_{\ell=k-K}^{k}\beta_\ell+\epsilon B.$$

Therefore, using $\sum_{l=0}^{\infty}\beta_l<\infty$, we obtain

$$\limsup_{k\to\infty}\sum_{\ell=0}^{k}\beta_{k-\ell}\gamma_\ell\leq\epsilon B.$$

Since ϵ is arbitrary, we conclude that $\limsup_{k\to\infty}\sum_{\ell=0}^{k}\beta_{k-\ell}\gamma_\ell=0$, implying

$$\lim_{k\to\infty}\sum_{\ell=0}^{k}\beta_{k-\ell}\gamma_\ell=0.$$

Suppose now $\sum_k\gamma_k<\infty$. Then, for any integer $M\geq 1$, we have

$$\sum_{k=0}^{M}\left(\sum_{\ell=0}^{k}\beta_{k-\ell}\gamma_\ell\right)=\sum_{\ell=0}^{M}\gamma_\ell\sum_{t=0}^{M-\ell}\beta_t\leq\sum_{\ell=0}^{M}\gamma_\ell B,$$

implying that

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

□

Proof of Lemma 8 Using the definition of projection error in Eq. (5), we have

$$e_i(k)=x_i(k+1)-v_i(k)+\alpha(k)d_i(k).$$

Taking the norms of both sides and using subgradient boundedness, we obtain

$$\|e_i(k)\|\leq\|x_i(k+1)-v_i(k)\|+\alpha(k)L.$$

Since $v_i(k) = \sum_{j=1}^m a_{ij}(k)x_j(k)$, the weight vector $a_i(k)$ is stochastic, and $x_j(k) \in X_j = X$ (cf. Assumption 6), it follows that $v_i(k) \in X$ for all i. Using the nonexpansive property of projection operation [cf. Eq. (35)] in the preceding relation, we obtain

$$\|e_i(k)\| \leq \|v_i(k) - \alpha(k)d_i(k) - v_i(k)\| + \alpha(k)L \leq 2\alpha(k)L,$$

completing the proof. □

Proof of Proposition 4 From Lemma 6, we have the following for all i and $k \geq 2$,

$$\begin{aligned}\|x_i(k) - y(k)\| \leq{}& m\rho(k-1,0)\sum_{j=1}^m \|x_j(0)\| + mL\sum_{r=0}^{k-2}\rho(k-1,r+1)\alpha(r) + 2\alpha(k-1)L \\ &+ \sum_{r=0}^{k-2}\rho(k-1,r+1)\sum_{j=1}^m \|e_j(r)\| + \|e_i(k-1)\| + \frac{1}{m}\sum_{j=1}^m \|e_j(k-1)\|.\end{aligned}$$

Using the upper bound on the projection error from Lemma 8, $\|e_i(k)\| \leq 2\alpha(k)L$ for all i and k, this can be rewritten as

$$\begin{aligned}\|x_i(k) - y(k)\| \leq m\rho(k-1,0)\sum_{j=1}^m \|x_j(0)\| &+ 3mL\sum_{r=0}^{k-2}\rho(k-1,r+1)\alpha(r) \\ &+ 6\alpha(k-1)L.\end{aligned} \tag{49}$$

Under Assumption 5 on the stepsize sequence, Proposition 2 implies the following bound for the disagreement metric $\rho(k,s)$: for all $k \geq s \geq 0$, $E[\rho(k,s)] \leq \beta(s)e^{-\mu\sqrt{k-s}}$, where μ is a positive scalar and $\beta(s)$ is an increasing sequence such that

$$\beta(s) \leq s^q \qquad \text{for all } q > 0 \text{ and all } s \geq S(q), \tag{50}$$

for some integer $S(q)$, i.e., for all $q > 0$, $\beta(s)$ is bounded by a polynomial s^q for sufficiently large s (where the threshold on s, $S(q)$, depends on q). Taking the expectation in Eq. (49) and using the preceding estimate on $\rho(k,s)$, we obtain

$$\begin{aligned}E[\|x_i(k) - y(k)\|] \leq m\beta(0)e^{-\mu\sqrt{k-1}}\sum_{j=1}^m \|x_j(0)\| &+ 3mL\sum_{r=0}^{k-2}\beta(r+1)e^{-\mu\sqrt{k-r-2}}\alpha(r) \\ &+ 6\alpha(k-1)L.\end{aligned}$$

We can bound $\beta(0)$ by $\beta(0) \leq S(1)$ by using Eq. (50) with $q = 1$ and the fact that β is an increasing sequence. Therefore, by taking the limit superior in the preceding relation and using $\alpha(k) \to 0$ as $k \to \infty$, we have for all i,

$$\limsup_{k\to\infty} E[\|x_i(k) - y(k)\|] \leq 3mL\sum_{r=0}^{k-2}\beta(r+1)e^{-\mu\sqrt{k-r-2}}\alpha(r).$$

Finally, note that $\lim_{k\to\infty}\beta(k+1)\alpha(k) \leq \lim_{k\to\infty}(k+1)\alpha(k) = 0$, where the inequality holds by using Eq. (50) with $q = 1$ and the equality holds by Assumption 5 on the stepsize. Since we also have $\sum_{k=0}^\infty e^{-\mu\sqrt{k}} < \infty$, Lemma 7 applies implying that

$$\lim_{k\to\infty}\sum_{r=0}^{k-2}\beta(r+1)e^{-\mu\sqrt{k-r-2}}\alpha(r) = 0.$$

Combining the preceding relations, we have $\lim_{k\to\infty} E[\|x_i(k) - y(k)\|] = 0$. Using Fatou's Lemma (which applies since the random variables $\|y(k) - x_i(k)\|$ are nonnegative for all i and k), we obtain

$$0 \leq E\Big[\liminf_{k\to\infty}\|y(k) - x_i(k)\|\Big] \leq \liminf_{k\to\infty} E[\|y(k) - x_i(k)\|] \leq 0.$$

Thus, the nonnegative random variable $\liminf_{k\to\infty} \|y(k) - x_i(k)\|$ has expectation 0, which implies that $\liminf_{k\to\infty} \|y(k) - x_i(k)\| = 0$ with probability one. □

Proof of Proposition 51 From Lemma 6, we have

$$\begin{aligned}\|x_i(k) - y(k)\| \le\ & m\rho(k-1,0)\sum_{j=1}^{m}\|x_j(0)\| + mL\sum_{r=0}^{k-2}\rho(k-1,r+1)\alpha(r) + 2\alpha(k-1)L \\ & + \sum_{r=0}^{k-2}\rho(k-1,r+1)\sum_{j=1}^{m}\|e_j(r)\| + \|e_i(k-1)\| + \frac{1}{m}\sum_{j=1}^{m}\|e_j(k-1)\|.\end{aligned}$$

Taking the expectation of both sides and using the estimate for the disagreement metric $\rho(k,s)$ from Proposition 3, i.e., for all $k \ge s \ge 0$, $E[\rho(k,s)] \le \kappa e^{-\mu(k-s)}$, for some scalars $\kappa, \mu > 0$, we obtain

$$\begin{aligned}E[\|x_i(k) - y(k)\|] \le\ & m\kappa e^{-\mu(k-1)}\sum_{j=1}^{m}\|x_j(0)\| + mL\kappa\sum_{r=0}^{k-2}e^{-\mu(k-r-2)}\alpha(r) + 2\alpha(k-1)L \\ & +\kappa\sum_{r=0}^{k-2}e^{-\mu(k-r-2)}\sum_{j=1}^{m}\|e_j(r)\| + \|e_i(k-1)\| + \frac{1}{m}\sum_{j=1}^{m}\|e_j(k-1)\|.\end{aligned}$$

By taking the limit superior in the preceding relation and using the facts that $\alpha(k) \to 0$, and $\|e_i(k)\| \to 0$ for all i as $k \to \infty$ (cf. Lemma 10(b)), we have for all i,

$$\limsup_{k\to\infty} E[\|x_i(k) - y(k)\|] \le mL\kappa\sum_{r=0}^{k-2}e^{-\mu(k-r-2)}\alpha(r)\ + \kappa\sum_{r=0}^{k-2}e^{-\mu(k-r-2)}\sum_{j=1}^{m}\|e_j(r)\|.$$

Finally, since $\sum_{k=0}^{\infty} e^{-\mu k} < \infty$ and both $\alpha(k) \to 0$ and $\|e_i(k)\| \to 0$ for all i, by Lemma 7, we have

$$\lim_{k\to\infty}\sum_{r=0}^{k-2}e^{-\mu(k-r-2)}\alpha(r) = 0 \quad \text{and} \quad \lim_{k\to\infty}\sum_{r=0}^{k-2}e^{-\mu(k-r-2)}\sum_{j=1}^{m}\|e_j(r)\| = 0.$$

Combining the preceding two relations, we have $\lim_{k\to\infty} E[\|x_i(k) - y(k)\|] = 0$. The second part of proposition follows using Fatou's Lemma and a similar argument used in the proof of Proposition 4. □

Proof of Proposition 6 This proof uses the result that

$$\liminf_{k\to\infty}\|x_i(k) - y(k)\| = 0 \quad \text{with probability one,} \tag{51}$$

which is derived using the same steps as in Proposition 4, but replacing the bound on the disagreement metric $\rho(k,s)$ from Proposition 2 with the one from Proposition 3.

Using the iterations (4) and (28), we obtain for all $k \ge 1$ and i,

$$\begin{aligned}y(k+1) - x_i(k+1) =\ & \Big(y(k) - \sum_{j=1}^{m}a_{ij}(k)x_j(k)\Big) - \alpha(k)\Big(\frac{1}{m}\sum_{j=1}^{m}d_j(k) - d_i(k)\Big) \\ & + \Big(\frac{1}{m}\sum_{j=1}^{m}e_j(k) - e_i(k)\Big).\end{aligned}$$

Using the doubly stochasticity of the weights $a_{ij}(k)$ and the subgradient boundedness (which holds by Assumption 7), this implies that

$$\sum_{i=1}^{m}\|y(k+1) - x_i(k+1)\| \le \sum_{i=1}^{m}\|y(k) - x_i(k)\| + 2Lm\alpha(k) + 2\sum_{i=1}^{m}\|e_i(k)\|. \tag{52}$$

Since $\alpha(k) \to 0$, it follows from Lemma 10(b) that $\|e_i(k)\| \to 0$ for all i. Eq. (52) then yields

$$\begin{aligned}\limsup_{k\to\infty} \sum_{i=1}^{m} \|y(k+1) - x_i(k+1)\| &\leq \liminf_{k\to\infty} \sum_{i=1}^{m} \|y(k) - x_i(k)\| \\ &\quad + \lim_{k\to\infty} \Big(2Lm\alpha(k) + 2\sum_{i=1}^{m} \|e_i(k)\|\Big) \\ &= \liminf_{k\to\infty} \sum_{i=1}^{m} \|y(k) - x_i(k)\|.\end{aligned}$$

Using $x_i(k) \in X_i$ for all i and k, it follows from Assumption 7 that the sequence $\{x_i(k)\}$ is bounded for all i. Therefore, the sequence $\{y(k)\}$ [defined by $y(k) = \frac{1}{m}\sum_{i=1}^{m} x_i(k)$, see Eq. (27)], and also the sequences $\|y(k) - x_i(k)\|$ are bounded. Combined with the preceding relation, this implies that the scalar sequence $\sum_{i=1}^{m} \|y(k) - x_i(k)\|$ is convergent.

By Eq. (51), we have $\liminf_{k\to\infty} \|x_i(k) - y(k)\| = 0$ with probability one. Since $\sum_{i=1}^{m} \|y(k) - x_i(k)\|$ converges, this implies that for all i, $\lim_{k\to\infty} \|x_i(k) - y(k)\| = 0$ with probability one, completing the proof. □

C Appendix - Simulation Analysis

In this section, we show via simulation that the choice of stepsizes plays a big role in the performance of the proposed algorithm. We also use simulation to understand the sensitivity of the algorithm's performance to the fading constant C from Eq. (6), which determines how fast communication failures increase as the agents' estimates move away from each other.

The results below concern a two-dimensional optimization problem, solved by five agents. Agent 1's objective function is $x^2 + y^2$, agent two's is $-x$, agent three's is x, agent four's is y and agent five's is $-y$. The joint objective is, therefore, $x^2 + y^2$ and the optimal solution is $(x, y) = (0, 0)$. The agents start from uniformly random locations in $[-5, 5]^2$. We assume that at every period at most one pair of agents communicates and the maximum communication probability is $\delta = 0.1$ and the fading constant $C = 2$ (see Eq. (6)) .

We consider two choices of stepsizes: $\alpha(k) = \frac{1}{k\log(k)}$ for all $k \geq 2$ and $\alpha(k) = \frac{200}{k\log(k)}$ for all k for all $k \geq 200$. In both cases, we use 0.1 for small k. Figures 2 and 3 show examples of sample paths using the two choices of stepsizes.

The second stepsize performed significantly better than the first. The algorithm converged to within 0.01 of the optimal value in 10000 iterations in 52% of the runs using the first stepsize, while it converged to within 10000 iterations in 98% of the runs using the second stepsize. The sample path graphs highlight that the dynamics of the agents' estimates is quite different using the two choices of stepsizes: in the first case, the agents first converged on a location and then slowly moved towards the optimum approximately together; in the second case, the agents quickly approached the region near the optimal solution but they took much longer to coalesce together. The dynamics with the second stepsize choice are explained by the fact that the stepsize is constant for the first 200 iterations, a period during which the agents can quickly move towards the region around the optimal solution.

We also performed a sensitivity analysis with respect to the constant C from Eq. (6). The objective is to understand how the environment, as measured by the fading constant C, impacts the performance of this multi-agent distributed algorithm. We consider the same simulation as above, using the first stepsize, but vary C. As C changes from 1 to 3, the percent of simulation runs that ended with convergence within 10000 iterations falls from 76% to 40% (see Figure 4). This suggests the decline in performance of the algorithm as C grows is not as sharp as one might expect, in particular, if the agents' initial position is not too far from the optimal solution.

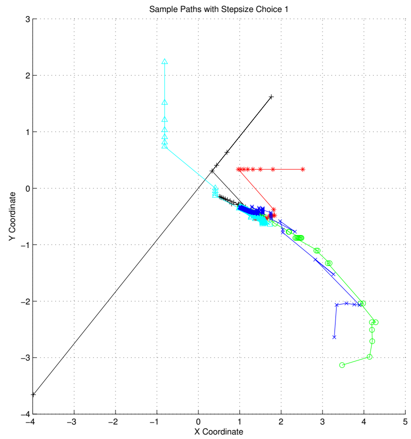

Fig. 2 Example of a sample path with stepsize choice of $\alpha(k) = \frac{1}{k\log(k)}$.

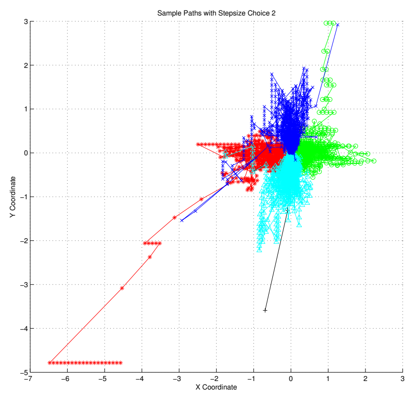

Fig. 3 Example of a sample path with stepsize choice of $\alpha(k) = \frac{200}{k\log(k)}$, for $k \geq 200$ and $\alpha(k) = 0.1$ for $k < 200$.

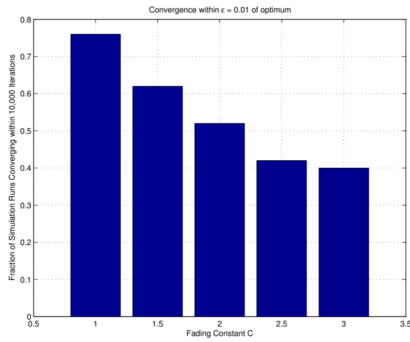

Fig. 4 Impact of fading constant C on probability of convergence within 10,000 iterations.