4. Supplementary File

4.1. Proof of Lemma 1

Proof. Let $\mathbf{x} \in \mathbb{R}^{nN}$ and $g_i \in \partial\rho_i(x_i)$ for all $i \in \mathcal{N}$. From convexity of ρ_i and Cauchy-Schwarz, it follows that $\rho_i(x_i) \leq \rho(\bar{x}_i) + \|g_i\|_2\|x_i - \bar{x}_i\|_2$ for all $i \in \mathcal{N}$. Hence, we have

$$\lambda\bar{\rho}(\mathbf{x}) + f(\mathbf{x}) \leq \lambda\rho(\bar{\mathbf{x}}) + f(\bar{\mathbf{x}}) + \sum_{i\in\mathcal{N}}\left(\lambda B_i\|x_i - \bar{x}_i\|_2 + \nabla_{x_i}f(\bar{\mathbf{x}})^\mathsf{T}(x_i - \bar{x}_i) + \frac{L_i}{2}\|x_i - \bar{x}_i\|_2^2\right).$$

Minimizing on both sides and using the separability of the right side, we have $\min_{\mathbf{x}\in\mathbb{R}^{nN}} \lambda\bar{\rho}(\mathbf{x}) + f(\mathbf{x}) \leq \lambda\bar{\rho}(\bar{\mathbf{x}}) + f(\bar{\mathbf{x}}) + \sum_{i\in\mathcal{N}} \min_{x_i\in\mathbb{R}^n} h_i(x_i)$, where $h_i(x_i) := \nabla_{x_i}f(\bar{\mathbf{x}})^\mathsf{T}(x_i - \bar{x}_i) + \lambda B_i\|x_i - \bar{x}_i\|_2 + \frac{L_i}{2}\|x_i - \bar{x}_i\|_2^2$. Let $\bar{x}_i^* := \operatorname{argmin}_{x_i\in\mathbb{R}^n} h_i(x_i)$. Then the first-order optimality conditions imply that $0 \in \nabla_{x_i}f(\bar{\mathbf{x}}) + L_i(\bar{x}_i^* - \bar{x}_i) + \lambda B_i\,\partial\|x_i - \bar{x}_i\|_2\Big|_{x_i=\bar{x}_i^*}$ for all $i \in \mathcal{N}$.

Let $\mathcal{I} := \{i \in \mathcal{N} :\ \|\nabla_{x_i}f(\bar{\mathbf{x}})\|_2 \leq \lambda B_i\}$. For each $i \in \mathcal{N}$, there are two possibilities.

Case 1: Suppose that $i \in \mathcal{I}$, i.e., $\|\nabla_{x_i}f(\bar{\mathbf{x}})\|_2 \leq \lambda B_i$. Since $\min_{x_i\in\mathbb{R}^n} h_i(x_i)$ has a unique solution, and $-\nabla_{x_i}f(\bar{\mathbf{x}}) \in \lambda B_i\,\partial\|x_i - \bar{x}_i\|_2\Big|_{x_i=\bar{x}_i}$ when $\|\nabla_{x_i}f(\bar{\mathbf{x}})\|_2 \leq \lambda B_i$, it follows that $\bar{x}_i^* = \bar{x}_i$ if and only if $\|\nabla_{x_i}f(\bar{\mathbf{x}})\|_2 \leq \lambda B_i$. Hence, $h_i(\bar{x}_i^*) = 0$.

Case 2: Suppose that $i \in \mathcal{I}^c := \mathcal{N} \setminus \mathcal{I}$, i.e., $\|\nabla_{x_i}f(\bar{\mathbf{x}})\|_2 > \lambda B_i$. In this case, $\bar{x}_i^* \neq \bar{x}_i$. From the first-order optimality condition, we have $\nabla_{x_i}f(\bar{\mathbf{x}}) + L_i(\bar{x}_i^* - \bar{x}_i) + \lambda B_i\frac{\bar{x}_i^* - \bar{x}_i}{\|\bar{x}_i^* - \bar{x}_i\|_2} = 0$. Let $s_i := \frac{\bar{x}_i^* - \bar{x}_i}{\|\bar{x}_i^* - \bar{x}_i\|_2}$ and $t_i := \|\bar{x}_i^* - \bar{x}_i\|_2$, then $s_i = \frac{-\nabla_{x_i}f(\bar{\mathbf{x}})}{L_i t_i + \lambda B_i}$. Since $\|s_i\|_2 = 1$, it follows that $t_i = \frac{\|\nabla_{x_i}f(\bar{\mathbf{x}})\|_2 - \lambda B_i}{L_i} > 0$, and $s_i = \frac{-\nabla_{x_i}f(\bar{\mathbf{x}})}{\|\nabla_{x_i}f(\bar{\mathbf{x}})\|_2}$. Hence, $\bar{x}_i^* = \bar{x}_i - \frac{\|\nabla_{x_i}f(\bar{\mathbf{x}})\|_2 - \lambda B_i}{L_i}\frac{\nabla_{x_i}f(\bar{\mathbf{x}})}{\|\nabla_{x_i}f(\bar{\mathbf{x}})\|_2}$, and $h_i(\bar{x}_i^*) = -\frac{(\|\nabla_{x_i}f(\bar{\mathbf{x}})\|_2 - \lambda B_i)^2}{2L_i}$.

From the α-optimality of $\bar{\mathbf{x}}$, it follows that

$$\sum_{i\in\mathcal{I}}\frac{(\|\nabla_{x_i}f(\bar{\mathbf{x}})\|_2 - \lambda B_i)^2}{2L_i} = -\sum_{i\in\mathcal{I}} h_i(\bar{x}_i^*) \leq \lambda\bar{\rho}(\bar{\mathbf{x}}) + f(\bar{\mathbf{x}}) - \min_{\mathbf{x}\in\mathbb{R}^{nN}} \lambda\bar{\rho}(\mathbf{x}) + f(\mathbf{x}) \leq \alpha,$$

which implies that $\|\nabla_{x_i}f(\bar{\mathbf{x}})\|_2 \leq \sqrt{2L_i\alpha} + \lambda B_i$ for all $i \in \mathcal{I}$. Moreover, $\|\nabla_{x_i}f(\bar{\mathbf{x}})\|_2 \leq \lambda B_i$ for all $i \in \mathcal{I}^c$. Hence, the result follows from these two inequalities. □

4.2. Proof of Lemma 2

Proof. For all $i \in \mathcal{N}$, since $\nabla\gamma_i$ is Lipschitz continuous with constant L_{γ_i}, for any $\mathbf{x}, \bar{\mathbf{x}} \in \mathbb{R}^{nN}$, we have $\gamma_i(x_i) \leq \gamma_i(\bar{x}_i) + \nabla\gamma_i(\bar{x}_i)^\mathsf{T}(x_i - \bar{x}_i) + \frac{L_{\gamma_i}}{2}\|x_i - \bar{x}_i\|_2^2$. Then, it follows that

$$\begin{aligned}\bar{\gamma}(\mathbf{x}) &\leq \sum_{i=1}^N \gamma_i(\bar{x}_i) + \nabla\gamma_i(\bar{x}_i)^\mathsf{T}(x_i - \bar{x}_i) + \frac{L_{\gamma_i}}{2}\|x_i - \bar{x}_i\|_2^2 \\ &\leq \bar{\gamma}(\bar{\mathbf{x}}) + \nabla\bar{\gamma}(\bar{\mathbf{x}})^\mathsf{T}(\mathbf{x} - \bar{\mathbf{x}}) + \sum_{i=1}^N \frac{L_{\gamma_i}}{2}\|x_i - \bar{x}_i\|_2^2. \end{aligned} \tag{15}$$

Let $h^{(k)}(\mathbf{x}) = \frac{1}{2}\|A\mathbf{x} - b - \lambda^{(k)}\theta^{(k)}\|_2^2$. It follows that $\nabla h^{(k)}$ is Lipschitz continuous with constant $\sigma_{\max}^2(A)$. Since $f^{(k)} = \lambda^{(k)}\bar{\gamma} + h^{(k)}$, the result follows from (15). □

4.3. Proof of Lemma 3

Proof. Fix $k \geq 1$. Suppose that $\mathbf{x}^{(k)}$ satisfies (9)(a). Then Lemma 1 implies that for all $i \in \mathcal{N}$

$$\|\nabla_{x_i}f^{(k)}(\mathbf{x}^{(k)})\|_2 = \|\lambda^{(k)}\nabla\gamma_i(x_i^{(k)}) + A_i^\mathsf{T}(A\mathbf{x}^{(k)} - b - \lambda^{(k)}\theta^{(k)})\|_2 \leq \sqrt{2L_i^{(k)}\alpha^{(k)}} + \lambda^{(k)}B_i.$$

Now, suppose that $\mathbf{x}^{(k)}$ satisfies (9)(b). Then triangular inequality immediately implies that $\|\nabla_{x_i} f^{(k)}(\mathbf{x}^{(k)})\|_2 \leq \xi^{(k)}/\sqrt{N} + \lambda^{(k)} B_i$ for all $i \in \mathcal{N}$. Combining the two inequalities, and further using triangular Cauchy-Schwarz inequalities, it follows for all $i \in \mathcal{N}$ that $\|A\mathbf{x}^{(k)} - b - \lambda^{(k)}\theta^{(k)}\|_2 \leq \frac{\max\left\{\sqrt{2L_i^{(k)}\alpha_k},\ \xi^{(k)}/\sqrt{N}\right\} + \lambda^{(k)}\left(B_i + \|\nabla\gamma(x_i^{(k)})\|_2\right)}{\sigma_{\min}(A_i)}$. Hence, we conclude by diving the above inequality by $\lambda^{(k)}$ and using the definition of $\theta^{(k+1)}$. □

4.4. Proof of Theorem 1

Proof. Let $A = [A_1, A_2, \ldots, A_N] \in \mathbb{R}^{m\times nN}$ such that $A_i \in \mathbb{R}^{m\times n}$ for all $i \in \mathcal{N}$. Throughout the proof we assume that $\sigma_{\max}(A) \geq \sqrt{\max_{i\in\mathcal{N}} d_i + 1}$, and $\sigma_{\min}(A_i) = \sqrt{d_i} \geq 1$ for all $i \in \mathcal{N}$, where $d_i \geq 1$ is the degree of $i \in \mathcal{N}$. Indeed, when A is chosen as described in Section 2.2.3 corresponding to graph $\mathcal{G}$, recall that we showed $\sigma^2_{\max}(A) = \psi_1$, where ψ_1 is the largest eigenvalue of the Laplacian Ω corresponding to $\mathcal{G}$. It is shown in (Grone & Merris, 1994) that when $\mathcal{G}$ is connected, one has $\psi_1 \geq \max_{i\in\mathcal{N}} d_i + 1 > 1$. Hence, $\sigma_{\max}(A) \geq \sqrt{\max_{i\in\mathcal{N}} d_i + 1} > 1$. Moreover, for A chosen as described in Section 2.2.3 corresponding to graph $\mathcal{G}$, again recall that $\sigma_{\min}(A_i) = \sqrt{d_i}$ for all $i \in \mathcal{N}$.

To keep notation simple, without loss of generality, we assume that $\underline{\gamma_i} = 0$ for all $i \in \mathcal{N}$. Hence, $\bar{\gamma}(\mathbf{x}) \geq 0$ for all $\mathbf{x} \in \mathbb{R}^{nN}$. Let $\mathbf{x}^*$ be a minimizer of (6). By Lipschitz continuity of $\nabla\gamma_i$, we have for all $i \in \mathcal{N}$

$$\|\nabla\gamma(x_i)\|_2 \leq L_{\gamma_i}\|x_i - x_i^*\|_2 + \|\nabla\gamma_i(x_i^*)\|_2. \tag{16}$$

We prove the theorem using induction. We show that, for an appropriately chosen bound R, $\|\mathbf{x}^{(k)} - \mathbf{x}^*\|_2 \leq R$ implies that $\|\mathbf{x}^{(k+1)} - \mathbf{x}^*\|_2 \leq R$, for all $k \geq 1$. Fix $k \geq 1$. First, suppose that $\mathbf{x}^{(k+1)}$ satisfies (9)(a), i.e. $P^{(k+1)}(\mathbf{x}^{(k+1)}) \leq P^{(k+1)}(\mathbf{x}^*) + \alpha^{(k+1)}$. By dividing both sides by $\lambda^{(k+1)}$, it follows from Assumption 1, $A\mathbf{x}^* = b$, and $f^{(k+1)}(\cdot) \geq 0$ that

$$\bar{\tau}\|\mathbf{x}^{(k+1)}\|_2 \leq \bar{\rho}(\mathbf{x}^*) + \bar{\gamma}(\mathbf{x}^*) + \frac{\lambda^{(k+1)}}{2}\left(\|\theta^{(k+1)}\|_2^2 + \frac{\alpha^{(k+1)}}{\left(\lambda^{(k+1)}\right)^2}\right). \tag{17}$$

Next, suppose $\mathbf{x}^{(k+1)}$ satisfies (9)(b). It follows from convexity of $P^{(k+1)}$ and Cauchy-Schwarz inequality that $P^{(k+1)}(\mathbf{x}^{(k+1)}) \leq P^{(k+1)}(\mathbf{x}^*) + \xi^{(k+1)}\|\mathbf{x}^{(k+1)} - \mathbf{x}^*\|_2$. Again, dividing both sides by $\lambda^{(k+1)}$, we get

$$\bar{\tau}\|\mathbf{x}^{(k+1)}\|_2 \leq \bar{\rho}(\mathbf{x}^*) + \bar{\gamma}(\mathbf{x}^*) + \frac{\lambda^{(k+1)}}{2}\|\theta^{(k+1)}\|_2^2 + \frac{\xi^{(k+1)}}{\lambda^{(k+1)}}\|\mathbf{x}^{(k+1)} - \mathbf{x}^*\|_2. \tag{18}$$

Combining the bounds for both cases, (17) and (18), and using triangular inequality, we have

$$\left(\bar{\tau} - \frac{\xi^{(k+1)}}{\lambda^{(k+1)}}\right)\|\mathbf{x}^{(k+1)} - \mathbf{x}^*\|_2 \leq \bar{F}^* + \bar{\tau}\|\mathbf{x}^*\|_2 + \frac{\lambda^{(k+1)}}{2}\left(\|\theta^{(k+1)}\|_2^2 + \frac{\alpha^{(k+1)}}{\left(\lambda^{(k+1)}\right)^2}\right), \tag{19}$$

for all $k \geq 0$. Note that $\{\lambda^{(k)}, \alpha^{(k)}, \xi^{(k)}\}$ is chosen in **DFAL** such that $\frac{\alpha^{(k)}}{(\lambda^{(k)})^2} = \frac{\alpha^{(1)}}{(\lambda^{(1)})^2}$ for all $k > 1$, and both $\frac{\xi^{(k)}}{\lambda^{(k)}} \searrow 0$ and $\lambda^{(k)} \searrow 0$ monotonically. Since $\sigma_{\min}(A_i) \geq 1$ for all $i \in \mathcal{N}$, the inductive assumption $\|\mathbf{x}^{(k)} - \mathbf{x}^*\|_2 \leq R$, (16), and Lemma 3 together imply that

$$\|\theta^{(k+1)}\|_2 \leq \min_{i\in\mathcal{N}}\left\{\max\left\{\sqrt{2L_i^{(1)}\frac{\alpha^{(1)}}{(\lambda^{(1)})^2}},\ \frac{\xi^{(1)}}{\lambda^{(1)}}\right\} + B_i + \|\nabla\gamma_i(x_i^*)\|_2 + L_{\gamma_i}R\right\}. \tag{20}$$

To simplify bounds further, choose $\alpha^{(1)} = \frac{1}{4N}\left(\lambda^{(1)}\bar{\tau}\right)^2$, and $\xi^{(1)} = \frac{1}{2}\lambda^{(1)}\bar{\tau}$ for $\lambda^{(1)} \leq \sigma^2_{\max}(A)/\bar{L}$, where $\bar{L} = \max_{i\in\mathcal{N}}\{L_{\gamma_i}\}$. Let $\bar{B} := \max_{i\in\mathcal{N}} B_i$ and $\bar{G} := \max\{\|\nabla\gamma_i(x_i^*)\|_2 :\ i \in \mathcal{N}\}$. Together with (19), (20) and $\sigma_{\max}(A) \geq 1$, this choice of parameters implies that

$$\frac{\bar{\tau}}{2}\|\mathbf{x}^{(k+1)} - \mathbf{x}^*\|_2 \leq \bar{F}^* + \bar{\tau}\|\mathbf{x}^*\|_2 + \frac{\lambda^{(1)}}{2}\left[\left(\frac{\bar{\tau}\sigma_{\max}(A)}{\sqrt{N}} + \bar{B} + \bar{G} + \bar{L}R\right)^2 + \frac{\bar{\tau}^2}{4N}\right].$$

Define $\beta_1 := \frac{2}{\bar{\tau}}\left(\bar{F}^* + \bar{\tau}\|\mathbf{x}^*\|_2\right)$, $\beta_2 := \frac{\bar{\tau}\sigma_{\max}(A)/\sqrt{N} + \bar{B} + \bar{G}}{\sqrt{\bar{\tau}}}$, $\beta_3 := \frac{\bar{L}}{\sqrt{\bar{\tau}}}$, and $\beta_4 := \frac{\bar{\tau}}{4N}$. Then we have that $\|\mathbf{x}^{(k+1)} - \mathbf{x}^*\|_2 \leq \beta_1 + \lambda^{(1)}\left[\left(\beta_2 + \beta_3 R\right)^2 + \beta_4\right]$.

Note that we are free to choose any $\lambda^{(1)} > 0$ satisfying $\lambda^{(1)} \leq \sigma^2_{\max}(A)/\bar{L}$. Our objective is to show that by appropriately choosing $\lambda^{(1)}$, we can guarantee that $\beta_1 + \lambda^{(1)} \left[\left(\beta_2 + \beta_3 R\right)^2 + \beta_4 \right] \leq R$, which would then complete the inductive proof. This is indeed true if the above quadratic inequality in R, has a solution, or equivalently if the discriminant

$$\Delta = (2\lambda^{(1)}\beta_2\beta_3 - 1)^2 - 4\lambda^{(1)}\beta_3^2 \left[\lambda^{(1)}(\beta_2^2 + \beta_4) + \beta_1\right]$$

is non-negative. Note that Δ is continuous in $\lambda^{(1)}$, and $\lim_{\lambda^{(1)} \to 0} \Delta = 1$. Thus, for all sufficiently small $\lambda^{(1)} > 0$, we have $\Delta \geq 0$. Hence, we can set $R = \frac{1 - 2\lambda^{(1)}\beta_2\beta_3 - \sqrt{\Delta}}{2\lambda^{(1)}{\beta_3}^2}$ for some $\lambda^{(1)} > 0$ such that $\Delta \geq 0$, and this will imply that $\|\mathbf{x}^{(k+1)} - \mathbf{x}^*\|_2 \leq R$ whenever $\|\mathbf{x}^{(k)} - \mathbf{x}^*\|_2 \leq R$ for all $k \geq 1$.

The induction will be complete if we can show that $\|\mathbf{x}^{(1)} - \mathbf{x}^*\|_2 \leq R$. Note that in **DFAL** we set $\theta^{(1)} = \mathbf{0}$. Hence, for $k = 0$, (19) implies that $\|\mathbf{x}^{(1)} - \mathbf{x}^*\|_2 \leq \beta_1 + \lambda^{(1)}\beta_4$. Hence, our choice of R guarantees that $\|\mathbf{x}^{(1)} - \mathbf{x}^*\|_2 \leq R$. This completes the induction.

Following the same arguments leading to (19), it can also be shown that for all $k \geq 0$

$$\left(\bar{\tau} - \frac{\xi^{(k+1)}}{\lambda^{(k+1)}}\right) \|\mathbf{x}_*^{(k+1)} - \mathbf{x}^*\|_2 \leq \bar{F}^* + \bar{\tau}\|\mathbf{x}^*\|_2 + \frac{\lambda^{(k+1)}}{2}\|\theta^{(k+1)}\|_2^2.$$

Therefore, we can conclude that $\|\mathbf{x}_*^{(k)} - \mathbf{x}^*\| \leq R$ for all $k \geq 1$ holds for the same R we selected above.

Note that Δ is a concave quadratic of $\lambda^{(1)}$ such that $\Delta = 1$ when $\lambda^{(1)} = 0$; hence, one of its roots is positive and the other one is negative. Moreover, $R \leq \frac{1}{2\lambda^{(1)}{\beta_3}^2} - \frac{\beta_2}{\beta_3}$ and the bound on R is decreasing in $\lambda^{(1)} > 0$. Hence, in order to get a smaller bound on R, we will choose $\lambda^{(1)}$ as the positive root of Δ. In particular, we set $\lambda^{(1)} = \frac{\sqrt{(\beta_2 + \beta_3\beta_1)^2 + \beta_4} - (\beta_2 + \beta_3\beta_1)}{2\beta_3\beta_4}$. □

4.5. Proof of Theorem 2

Proof. The proof directly follows from Theorem 3.3 in (Aybat & Iyengar, 2012). For the sake of completeness, we also provide the proof here. Let $\mathbf{x}^*$ denote an optimal solution to (6).

Note that (a) follows immediately from Cauchy-Schwarz and the definition of $\theta^{(k+1)}$.

$$\|A\mathbf{x}^{(k)} - b\|_2 \leq \|A\mathbf{x}^{(k)} - b - \lambda^{(k)}\theta^{(k)}\|_2 + \lambda^{(k)}\|\theta^{(k)}\|_2 = \lambda^{(k)}(\|\theta^{(k+1)}\|_2 + \|\theta^{(k)}\|_2) \leq 2B_\theta\lambda^{(k)}.$$

First, we prove the second inequality in (b). Suppose that $\mathbf{x}^{(k)}$ satisfies (9)(a), which implies that $\bar{F}(\mathbf{x}^{(k)}) + \frac{\lambda^{(k)}}{2}\|\theta^{(k+1)}\|_2^2 \leq \bar{F}(\mathbf{x}^*) + \frac{\lambda^{(k)}}{2}\|\theta^{(k)}\|_2^2 + \frac{\alpha^{(k)}}{\lambda^{(k)}}$. Now, suppose that $\mathbf{x}^{(k)}$ satisfies (9)(b). From the convexity of $P^{(k)}$ and Cauchy-Schwarz, it follows that $P^{(k)}(\mathbf{x}^{(k)}) \leq P^{(k)}(\mathbf{x}^*) + \xi^{(k)}\|\mathbf{x}^{(k)} - \mathbf{x}^*\|_2$. Hence, dividing it by $\lambda^{(k)}$, we have $\bar{F}(\mathbf{x}^{(k)}) + \frac{\lambda^{(k)}}{2}\|\theta^{(k+1)}\|_2^2 \leq \bar{F}(\mathbf{x}^*) + \frac{\lambda^{(k)}}{2}\|\theta^{(k)}\|_2^2 + \frac{\xi^{(k)}}{\lambda^{(k)}}$. Therefore, for all $k \geq 1$, $\mathbf{x}^{(k)}$ satisfies the second inequality in (b) since it also satisfies

$$\bar{F}(\mathbf{x}^{(k)}) - \bar{F}^* \leq \lambda^{(k)} \left(\frac{\|\theta^{(k)}\|_2^2 - \|\theta^{(k+1)}\|_2^2}{2} + \frac{\max\left\{\alpha^{(k)}, \xi^{(k)}\|\mathbf{x}^{(k)} - \mathbf{x}^*\|_2\right\}}{(\lambda^{(k)})^2} \right).$$

Now, in order to prove the first inequality in (b), we will exploit the primal-dual relations of the following two pairs of problems:

$$(\mathcal{P}) : \min_{\mathbf{x}\in\mathbb{R}^{nN}} \{\bar{F}(\mathbf{x}) : \ A\mathbf{x} = b\}, \qquad (\mathcal{D}) : \max_{\theta\in\mathbb{R}^m} b^\top\theta - \bar{F}^*(A^\top\theta),$$

$$(\mathcal{P}_k) : \min_{\mathbf{x}\in\mathbb{R}^{nN}} \lambda^{(k)}\bar{F}(\mathbf{x}) + \tfrac{1}{2}\|A\mathbf{x} - b_k\|_2^2, \quad (\mathcal{D}_k) : \max_{\theta\in\mathbb{R}^m} \lambda^{(k)}(b^\top\theta - \bar{F}^*(A^\top\theta)) - \tfrac{(\lambda^{(k)})^2}{2}h(\theta),$$

where $b_k := b + \lambda^{(k)}\theta^{(k)}$, $h(\theta) := \|\theta - \theta^{(k)}\|_2^2 - \|\theta^{(k)}\|_2^2$, and $\bar{F}^*$ denotes the convex conjugate of $\bar{F}$. Note that problem $(\mathcal{P}_k)$ is nothing but the subproblem in (7). Therefore, from weak-duality between $(\mathcal{P}_k)$ and $(\mathcal{D}_k)$, it follows that

$$P^{(k)}(\mathbf{x}^{(k)}) = \lambda^{(k)}\bar{F}(\mathbf{x}^{(k)}) + \tfrac{1}{2}\|A\mathbf{x}^{(k)} - b_k\|_2^2 \geq \lambda^{(k)}(b^\top\theta^* - \bar{F}^*(A^\top\theta^*)) - \frac{(\lambda^{(k)})^2}{2}h(\theta^*).$$

Note that from strong duality between $(\mathcal{P})$ and $(\mathcal{D})$, it follows that $\bar{F}^* = \bar{F}(\mathbf{x}^*) = b^\top\theta^* - \bar{F}^*(A^\top\theta^*)$. Therefore, dividing the above inequality by $\lambda^{(k)}$, we obtain

$$\bar{F}(\mathbf{x}^{(k)}) - \bar{F}^* \geq -\frac{\lambda^{(k)}}{2}\left(\|\theta^*\|_2^2 - 2(\theta^*)^\top\theta^{(k)} + \|\theta^{(k+1)}\|_2^2\right) \geq -\frac{\lambda^{(k)}}{2}\left(\|\theta^*\|_2 + B_\theta\right)^2.$$

□

4.6. Proof of Theorem 3

Proof. We assume that $\sigma_{\max}(A) \geq \sqrt{\max_{i\in\mathcal{N}} d_i + 1}$, and $\sigma_{\min}(A_i) = \sqrt{d_i} \geq 1$ for all $i \in \mathcal{N}$, where d_i denotes the degree of $i \in \mathcal{N}$. As discussed in the proof of Theorem 1, this is a valid assumption for distributed optimization problem in (4). Let θ^* denote an optimal dual solution to (6). Note that from the first-order optimality conditions for (6), we have $\mathbf{0} \in \nabla\gamma_i(x_i^*) + A_i^\top\theta^* + \partial\rho_i(x_i)|_{x_i=x_i^*}$; hence, $\|A_i^\top\theta^*\|_2 \leq B_i + G_i$. Therefore, $\|\theta^*\|_2 \leq \min_{i\in\mathcal{N}} \frac{B_i+G_i}{\sigma_{\min}(A_i)}$.

Given $0 < \lambda^{(1)} \leq \sigma_{\max}^2(A)/\bar{L}$, choose $\alpha^{(1)}, \xi^{(1)} > 0$ such that $\alpha^{(1)} = \frac{1}{4N}\left(\lambda^{(1)}\bar{\tau}\right)^2$, and $\xi^{(1)} = \frac{1}{2}\lambda^{(1)}\bar{\tau}$. Then Lemma 3 and $\sigma_{\max}(A) \geq 1$ together imply that for all $k \geq 1$

$$\|\theta^{(k)}\|_2 \leq \min_{i\in\mathcal{N}} \left\{ \frac{\bar{\tau}\sigma_{\max}(A)/\sqrt{N} + B_i + G_i}{\sigma_{\min}(A_i)} \right\} := B_\theta. \tag{21}$$

Hence, note that $\|\theta^*\|_2 \leq B_\theta$.

To simplify notation, suppose that $\lambda^{(1)} = \min\left\{1, \sigma_{\max}^2(A)/\bar{L}\right\} = 1$. (19) implies that for all $k \geq 1$

$$\|\mathbf{x}^{(k)} - \mathbf{x}^*\|_2 \leq \frac{2}{\bar{\tau}}\left[\bar{F}^* + \bar{\tau}\|x^*\|_2 + \tfrac{1}{2}\left(B_\theta^2 + \frac{\bar{\tau}^2}{4N}\right)\right] := B_x. \tag{22}$$

Note that (22) implies that $\frac{\xi^{(1)}}{(\lambda^{(1)})^2} B_x = \frac{1}{\lambda^{(1)}}\frac{\bar{\tau}}{2}B_x \geq \frac{1}{2}B_\theta^2 + \frac{\bar{\tau}^2}{8N} \geq \frac{5}{8N}\bar{\tau}^2 \geq \frac{\alpha^{(1)}}{(\lambda^{(1)})^2}$, where we used the fact $B_\theta \geq \frac{\sigma_{\max}(A)}{\max_{i\in\mathcal{N}}\{\sigma_{\min}(A_i)\}}\frac{\bar{\tau}}{\sqrt{N}} \geq \frac{\bar{\tau}}{\sqrt{N}}$. Note that the last inequality follows from our assumption on A stated at the beginning of the proof, i.e. $\sigma_{\max}(A) \geq \sqrt{\max_{i\in\mathcal{N}} d_i + 1}$ and $\sigma_{\min}(A_i) = d_i$ for all $i \in \mathcal{N}$. Hence, Theorem 2, $\lambda^{(1)} = 1$, and $\|\theta^*\|_2 \leq B_\theta$ imply that

$$N_{\mathbf{DFAL}}^f(\epsilon) \leq \log_{\frac{1}{c}}\left(\frac{2B_\theta}{\epsilon}\right) = \log_{\frac{1}{c}}\left(2\min_{i\in\mathcal{N}}\left\{\frac{\bar{\tau}\sigma_{\max}(A)/\sqrt{N} + B_i + G_i}{\sigma_{\min}(A_i)\epsilon}\right\}\right) := \bar{N}^f, \tag{23}$$

$$\begin{aligned} N_{\mathbf{DFAL}}^o(\epsilon) &\leq \log_{\frac{1}{c}}\left(\frac{1}{\epsilon}\ \max\left\{\tfrac{1}{2}\left(\|\theta^*\|_2 + B_\theta\right)^2,\ B_\theta^2 + \bar{F}^* + \bar{\tau}\|x^*\|_2 + \frac{\bar{\tau}^2}{8N}\right\}\right), \\ &= \log_{\frac{1}{c}}\left(\frac{2B_\theta^2 + \bar{F}^* + \bar{\tau}\|x^*\|_2 + \frac{\bar{\tau}^2}{8N}}{\epsilon}\right) := \bar{N}^o. \end{aligned} \tag{24}$$

Since $\alpha^{(1)} = \frac{1}{4N}\left(\lambda^{(1)}\bar{\tau}\right)^2$, we have $\sqrt{\alpha^{(k)}} = \frac{\bar{\tau}}{\sqrt{4N}}c^k$. Hence, Lemma 5 implies that

$$N^{(k)} \leq 2B_x\sqrt{\frac{2(\lambda^{(k)}\bar{L} + \sigma_{\max}^2(A))}{\alpha^{(k)}}} \leq \frac{8B_x\sqrt{N}}{\bar{\tau}}\sigma_{\max}(A)c^{-k}. \tag{25}$$

Hence, (23) and (25) imply that the total number of **MS-APG** iterations to compute an ϵ-feasible solution can be bounded above:

$$\begin{aligned} \sum_{k=1}^{N_{\mathbf{DFAL}}^f(\epsilon)} N^{(k)} &\leq \frac{8B_x\sqrt{N}}{\bar{\tau}}\sigma_{\max}(A)\sum_{k=1}^{\bar{N}^f} c^{-k} \leq \frac{8B_x\sqrt{N}}{c(1-c)\bar{\tau}}\sigma_{\max}(A)\left(\frac{1}{c}\right)^{\bar{N}^f}, \\ &\leq \frac{16B_x\sqrt{N}}{c(1-c)\bar{\tau}}\min_{i\in\mathcal{N}}\left\{\frac{\bar{\tau}\sigma_{\max}(A)/\sqrt{N} + B_i + G_i}{\sigma_{\min}(A_i)\epsilon}\right\}\frac{\sigma_{\max}(A)}{\epsilon} = \mathcal{O}\left(\frac{\sigma_{\max}^2(A)}{\min_{i\in\mathcal{N}}\sigma_{\min}(A_i)}\frac{1}{\epsilon}\right). \end{aligned}$$

Similarly, (24) and (25) imply that the total number of **MS-APG** iterations to compute an ϵ-optimal solution can be bounded above:

$$\sum_{k=1}^{N_{\mathbf{DFAL}}^o(\epsilon)} N^{(k)} \leq \frac{8B_x\sqrt{N}}{c(1-c)\bar{\tau}}\sigma_{\max}(A)\left(\frac{1}{c}\right)^{\bar{N}^0} = \mathcal{O}\left(\frac{\sigma_{\max}^3(A)}{\min_{i\in\mathcal{N}}\sigma_{\min}^2(A_i)}\frac{1}{\epsilon}\right). \tag{26}$$

□

4.7. Proof of Lemma 6

Proof. Given any convex function $\rho : \mathbb{R}^n \to \mathbb{R}$ and $\bar{x} \in \mathbb{R}^n$, in order to simplify the notation throughout the proof, $\partial\rho(x)|_{x=\bar{x}} \subset \mathbb{R}^n$, the subdifferential of ρ at $\bar{x}$, will be written as $\partial\rho(\bar{x})$. Given $\bar{x} \in \mathbb{R}^n$, there exists $\nu \in \partial P(\bar{x})$ such that $\|\nu\|_2 \leq \xi$, if and only if $\|\nu^*\| \leq \xi$, where $\nu^* = \mathrm{argmin}\{\|\nu\|_2 : \ \nu \in \partial P(\bar{x})\}$. Note that $\partial P(\bar{x}) = \lambda\partial\rho(\bar{x}) + \nabla f(\bar{x})$, and

$$\partial\rho(\bar{x}) = \beta_1 \prod_{k=1}^{K} \partial\|\bar{x}_{g(k)}\|_1 + \beta_2 \prod_{k=1}^{K} \partial\|\bar{x}_{g(k)}\|_2, \tag{27}$$

where $\prod$ denotes the Cartesian product. Since the groups $\{g(k)\}_{k=1}^K$ are not overlapping with each other, the minimization problem is separable in groups. Hence, for all $k \in [1, K]$, we have $\nu^*_{g(k)} = \pi^*_{g(k)} + \omega^*_{g(k)} + \nabla_{x_{g(k)}} f(\bar{x})$ such that

$$\begin{aligned} (\pi^*_{g(k)}, \omega^*_{g(k)}) \quad &= \mathrm{argmin}\, \|\pi_{g(k)} + \omega_{g(k)} + \nabla_{x_{g(k)}} f(\bar{x})\|_2^2 \\ &s.t. \quad \pi_{g(k)} \in \lambda\beta_1\partial\|\bar{x}_{g(k)}\|_1, \quad \omega_{g(k)} \in \lambda\beta_2\partial\|\bar{x}_{g(k)}\|_2. \end{aligned} \tag{28}$$

Fix $k \in [1, K]$. We will consider the solution to above problem in two cases. Suppose that $\bar{x}_{g(k)} = \mathbf{0}$. Since $\partial\|\mathbf{0}\|_1$ is the unit ℓ_∞-ball, and $\partial\|\mathbf{0}\|_2$ is the unit ℓ_2-ball, (28) can be equivalently written as

$$\begin{aligned} (\pi^*_{g(k)}, \omega^*_{g(k)}) \quad &= \mathrm{argmin}\, \|\pi_{g(k)} + \omega_{g(k)} + \nabla_{x_{g(k)}} f(\bar{x})\|_2^2 \\ &s.t. \quad \|\pi_{g(k)}\|_\infty \leq \lambda\beta_1, \quad \|\omega_{g(k)}\|_2 \leq \lambda\beta_2. \end{aligned} \tag{29}$$

Clearly, it follows from Euclidean projection on to ℓ_2-ball that

$$\omega^*_{g(k)} = -(\pi^*_{g(k)} + \nabla_{x_{g(k)}} f(\bar{x})) \min\left\{1, \frac{\lambda\beta_2}{\|\pi^*_{g(k)} + \nabla_{x_{g(k)}} f(\bar{x})\|_2}\right\}.$$

Hence, $\|\pi^*_{g(k)} + \omega^*_{g(k)} + \nabla_{x_{g(k)}} f(\bar{x})\|_2 = \max\{0, \ \|\pi^*_{g(k)} + \nabla_{x_{g(k)}} f(\bar{x})\|_2 - \lambda\beta_2\}$. Therefore,

$$\pi^*_{g(k)} = \mathrm{argmin}\{\|\pi_{g(k)} + \nabla_{x_{g(k)}} f(\bar{x})\|_2 : \ \|\pi_{g(k)}\|_\infty \leq \lambda\beta_1\} = -\,\mathrm{sgn}(\nabla_{x_{g(k)}} f(\bar{x})) \odot \min\{|\nabla_{x_{g(k)}} f(\bar{x})|, \lambda\beta_1\}.$$

Now, suppose that $\bar{x}_{g(k)} \neq \mathbf{0}$. This implies that $\partial\|\bar{x}_{g(k)}\|_2 = \{\bar{x}_{g(k)}/\|\bar{x}_{g(k)}\|_2\}$. Hence, when $\bar{x}_{g(k)} \neq \mathbf{0}$, we have $\omega^*_{g(k)} = \lambda\beta_2\bar{x}_{g(k)}/\|\bar{x}_{g(k)}\|_2$, and the structure of $\partial\|\cdot\|_1$ implies that $\pi^*_j = \lambda\beta_1\,\mathrm{sgn}\,(\bar{x}_j)$ for all $j \in g(k)$ such that $|\bar{x}_j| > 0$; and it follows from (28) that for all $j \in g(k)$ such that $\bar{x}_j = 0$, we have

$$\pi^*_j = \mathrm{argmin}\left\{\left(\pi_j + \tfrac{\partial}{\partial x_j} f(\bar{x})\right)^2 : \ |\pi_j| \leq \lambda\beta_1\right\} = -\,\mathrm{sgn}\left(\tfrac{\partial}{\partial x_j} f(\bar{x})\right) \min\left\{\left|\tfrac{\partial}{\partial x_j} f(\bar{x})\right|, \ \lambda\beta_1\right\}.$$

□

4.8. Proof of Lemma 7

Proof. Since the groups are not overlapping with each other, the proximal problem becomes separable in groups. Let $n_k := |g(k)|$ for all k. Thus, it suffices to show that $\min_{x_{g(k)} \in \mathbb{R}^{n_k}} \{\beta_1\|x\|_1 + \beta_2\|x_{g(k)}\|_2 + \frac{1}{2t}\|x_{g(k)} - \bar{x}_{g(k)}\|_2^2\}$ has a closed form solution as shown in the statement for some fixed k. By the definition of dual norm, we have

$$\min_{x_{g(k)} \in \mathbb{R}^{n_k}} \beta_1\|x_{g(k)}\|_1 + \beta_2\|x_{g(k)}\|_2 + \frac{1}{2t}\|x_{g(k)} - \bar{x}_{g(k)}\|_2^2, \tag{30}$$

$$= \min_{x_{g(k)} \in \mathbb{R}^{n_k}} \max_{\|u_1\|_\infty \leq \beta_1} u_1^\top x_{g(k)} + \max_{\|u_2\|_2 \leq \beta_2} u_2^\top x_{g(k)} + \frac{1}{2t}\|x_{g(k)} - \bar{x}_{g(k)}\|_2^2,$$

$$= \max_{\substack{\|u_1\|_\infty \leq \beta_1 \\ \|u_2\|_2 \leq \beta_2}} \min_{x \in \mathbb{R}^n} (u_1 + u_2)^\top x_{g(k)} + \frac{1}{2t}\|x_{g(k)} - \bar{x}_{g(k)}\|_2^2, \tag{31}$$

$$= \max_{\substack{\|u_1\|_\infty \leq \beta_1 \\ \|u_2\|_2 \leq \beta_2}} (u_1 + u_2)^\top \bar{x}_{g(k)} - \frac{t}{2}\|u_1 + u_2\|_2^2. \tag{32}$$

Let (u_1^*, u_2^*) be the optimal solution of (32). Since $x^p_{g(k)}$ is the optimal solution to (30), it follows from (31) that

$$x^p_{g(k)} = \bar{x}_{g(k)} - t(u_1^* + u_2^*). \tag{33}$$

Note that (32) can be equivalently written as $\min\{\|u_1 + u_2 - \frac{1}{t}\bar{x}_{g(k)}\|_2^2 : \ \|u_1\|_\infty \leq \beta_1, \|u_2\|_2 \leq \beta_2\}$. Minimizing over u_2, we have

$$u_2^*(u_1) = \left(\frac{1}{t}\bar{x}_{g(k)} - u_1\right) \min\left\{\frac{\beta_2}{\|\frac{1}{t}\bar{x}_{g(k)} - u_1\|_2}, 1\right\}. \tag{34}$$

Hence, we have

$$u_1^* = \underset{\|u_1\|_\infty \leq \beta_1}{\operatorname{argmin}} \left\|\left(u_1 - \frac{1}{t}\bar{x}_{g(k)}\right) \max\left\{1 - \frac{\beta_2}{\|u_1 - \frac{1}{t}\bar{x}_{g(k)}\|_2}, 0\right\}\right\|_2 = \underset{\|u_1\|_\infty \leq \beta_1}{\operatorname{argmin}} \max\{\|u_1 - \frac{1}{t}\bar{x}_{g(k)}\|_2 - \beta_2, \ 0\}.$$

Clearly, $u_1^* = \operatorname{argmin}_{\|u_1\|_\infty \leq \beta_1} \|(u_1 - \frac{1}{t}\bar{x}_{g(k)})\|_2 = \operatorname{sgn}(\bar{x}_{g(k)}) \min\left\{\frac{1}{t}|\bar{x}_{g(k)}|, \beta_1\right\}$. The final result follows from combining (33) and (34). □

4.9. Improved rate for asynchronous DFAL

Let $\mathcal{R}$ denote a discrete random variable uniformly distributed over the set $\mathcal{N}$. Let $[U_1, U_2, \ldots, U_N]$ denote a partition of the nN-dimensional identity matrix where $U_i \in \mathbb{R}^{nN \times n}$, $i = 1, \ldots, N$. In the rest, given $\mathbf{h} \in \mathbb{R}^{nN}$, we denote $\mathbf{h}_{[\mathcal{R}]} := U_\mathcal{R} U_\mathcal{R}^\top \mathbf{h}$. Consider the composite convex optimization problem

$$\Phi^* := \min_{\mathbf{y} \in \mathbb{R}^{nN}} \Phi(\mathbf{y}) := \sum_{i=1}^N \rho_i(y_i) + f(\mathbf{y}), \tag{35}$$

where $\rho_i : \mathbb{R}^n \to \mathbb{R}$ is a closed convex function for all $i \in \mathcal{N}$ such that $\mathbf{prox}_{t\rho_i}$ can be computed efficiently for all $t > 0$ and $i \in \mathcal{N}$, and $f : \mathbb{R}^{nN} \to \mathbb{R}$ is a differentiable convex function such that for some $\{L_i\}_{i \in \mathcal{N}} \subset \mathbb{R}_{++}$, f satisfies

$$\mathbb{E}[f(\mathbf{y} + \mathbf{h}_{[\mathcal{R}]})] \leq f(\mathbf{y}) + \frac{1}{N}\left(\langle \nabla f(\mathbf{y}), \mathbf{h}\rangle + \frac{1}{2}\sum_{i \in \mathcal{N}} L_i \|h_i\|_2^2\right) \tag{36}$$

for all $\mathbf{y}, \mathbf{h} \in \mathbb{R}^{nN}$. Fercoq & Richtárik (2013) proposed the accelerated proximal coordinate descent algorithm **ARBCD** (see Figure 6) to solve (35). They showed that for a given $\alpha > 0$, the iterate sequence $\{\mathbf{z}^{(\ell)}, \mathbf{u}^{(\ell)}\}$ computed by **ARBCD** satisfies

$$\mathbb{E}\left[\Phi\left(\left(\frac{1}{Nt^{(\ell)}}\right)^2 \mathbf{u}^{(\ell+1)} + \mathbf{z}^{(\ell+1)}\right) - \Phi^*\right] \leq \alpha, \quad \forall \ell \geq 2N\sqrt{\frac{C}{\alpha}}, \tag{37}$$

where

$$C := \min_{\mathbf{y}^* \in \mathcal{Y}^*} (1 - \tfrac{1}{N})\left(\Phi\left(\mathbf{z}^{(0)}\right) - \Phi^*\right) + \tfrac{1}{2}\sum_{i \in \mathcal{N}} L_i \|z_i^{(0)} - y_i^*\|_2^2, \tag{38}$$

and $\mathcal{Y}^*$ denotes the set of optimal solutions.

Algorithm ARBCD $(\mathbf{z}^{(0)})$

1: $\ell \leftarrow 0, \quad t^{(0)} \leftarrow 1, \quad u_i^{(1)} \leftarrow \mathbf{0}, \quad \forall i \in \mathcal{N}$
2: **while** $\ell \geq 0$ **do**
3: $\quad i$ is a sample of $\mathcal{R}$
4: $\quad z_i^{(\ell+1)} \leftarrow \mathbf{prox}_{t^{(\ell)}\rho_i/L_i}\left(z_i^{(\ell)} - \frac{t^{(\ell)}}{L_i}\nabla_{y_i} f\left(\left(\frac{1}{Nt^{(\ell)}}\right)^2 \mathbf{u}^{(\ell)} + \mathbf{z}^{(\ell)}\right)\right)$
5: $\quad u_i^{(\ell+1)} \leftarrow u_i^{(\ell)} + N^2 t^{(\ell)}(1 - t^{(\ell)})\left(z_i^{(\ell+1)} - z_i^{(\ell)}\right)$
6: $\quad z_{-i}^{(\ell+1)} \leftarrow z_{-i}^{(\ell)}, \quad u_{-i}^{(\ell+1)} \leftarrow u_{-i}^{(\ell)}$
7: $\quad t^{(\ell+1)} \leftarrow \frac{1+\sqrt{1+(2Nt^{(\ell)})^2}}{2N}$
8: **end while**

Figure 6. Accelerated Randomized Proximal Block Coordinate Descent (ARBCD) algorithm

In the following result, we establish that the bound (36) can be exploited for designing an accelerated version of asynchronous **DFAL**.

Lemma 8. *Fix $\alpha > 0$, and $p \in (0,1)$. Let $\{\mathbf{z}_k^{(\ell)}, \mathbf{u}_k^{(\ell)}\}_{\ell\in\mathbb{Z}_+}$, $k = 1,\ldots,K$, denote the iterate sequence corresponding to $K := \log(1/p)$ independent calls to* ***ARBCD***$(\mathbf{y}^{(0)})$*. Define $\mathbf{y}_k := \left(\frac{1}{Nt^{(T)}}\right)^2 \mathbf{u}_k^{(T+1)} + \mathbf{z}_k^{(T+1)}$ for $k = 1,\ldots,K$, and $T := 2N\sqrt{\frac{2C}{\alpha}}$. Then*

$$\mathbb{P}\left(\min_{k=1,\ldots,K} \Phi(\mathbf{y}_k) - \Phi^* \leq \alpha\right) \geq 1 - p.$$

Proof. Since the sequence $\{\mathbf{y}_k\}_{k=1}^K$ is i.i.d., and each $\mathbf{y}_k$ satisfies $\mathbb{E}[\Phi(\mathbf{y}_k) - \Phi^*] \leq \frac{\alpha}{2}$, Markov's inequality implies that $\mathbb{P}(\Phi(\mathbf{y}_k) - \Phi^* > \alpha) \leq \mathbf{E}[\Phi(\mathbf{y}_k) - \Phi^*]/\alpha \leq \frac{1}{2}$ for $1 \leq k \leq K$. Therefore, we have

$$\mathbb{P}\left(\min_{k=1,\ldots,K} \Phi(\mathbf{y}_k) - \Phi^* \leq \alpha\right) = 1 - \prod_{k=1}^{K} \mathbb{P}(\Phi(\mathbf{y}_k) - \Phi^* > \alpha) \leq \left(\frac{1}{2}\right)^K = 1 - p.$$

□

From Lemma 8 it follows that we can compute $\mathbf{y}_\alpha$ such that $\mathbb{P}\left(\Phi(\mathbf{y}_\alpha) - \Phi^* \leq \alpha\right) \geq 1 - p$ in at most $2N\sqrt{\frac{2C}{\alpha}}\log(\frac{1}{p})$ **ARBCD** iterations. This new oracle can be used to construct an asynchronous version of **DFAL** algorithm with $\mathcal{O}(1/\epsilon)$ complexity.

Theorem 4. *Fix $\epsilon > 0$ and $p \in (0,1)$. Consider a* asynchronous *variant of* ***DFAL*** *where (9)(a) in Figure 1 is replaced by*

$$\mathbb{P}\left(P^{(k)}\big(\mathbf{x}^{(k)}\big) - P^{(k)}\big(\mathbf{x}_*^{(k)}\big) \leq \alpha^{(k)}\right) \geq (1-p)^{\frac{1}{N(\epsilon)}}, \tag{39}$$

where $N(\epsilon) = \log_{\frac{1}{c}}\left(\frac{\bar{C}}{\epsilon}\right)$ is defined in Corollary 1. Then $\{x_i^{(N(\epsilon))}\}_{i\in\mathcal{N}}$, satisfies

$$\mathbf{P}(\epsilon) := \mathbb{P}\left(\Big|\sum_{i\in\mathcal{N}} F_i\left(x_i^{(N(\epsilon))}\right) - F^*\Big| \leq \epsilon, \ \text{ and } \max_{(i,j)\in\mathcal{E}}\left\{\|x_i^{(N(\epsilon))} - x_j^{(N(\epsilon))}\|_2\right\} \leq \epsilon\right) \geq 1 - p,$$

and $\mathcal{O}\left(\frac{1}{\epsilon}\log\left(\frac{1}{p}\right)\right)$ ***ARBCD*** *iterations are required to compute $\{x_i^{(N(\epsilon))}\}_{i\in\mathcal{N}}$.*

Proof. Consider the k-th **DFAL** subproblem $\min P^{(k)}(\mathbf{x}) := \lambda^{(k)} \sum_{i\in\mathcal{N}} \rho_i(x_i) + f^{(k)}(\mathbf{x})$, where $f^{(k)}$ is defined in (8). Let $\tilde{L}_i^{(k)} := \lambda^{(k)} L_{\gamma_i} + d_i$ for all $i \in \mathcal{N}$. Then it can be easily shown that $f^{(k)}$ satisfies (36) with constants $\{\tilde{L}_i^{(k)}\}_{i\in\mathcal{N}}$ for all $1 \leq k \leq N(\epsilon)$. Hence, **ARBCD** algorithm can be used to solve $\min P^{(k)}(\mathbf{x})$ with the iteration complexity given in Lemma 8. Consider the random event

$$\Delta := \bigcap_{k=1}^{N(\epsilon)} \left\{P^{(k)}(\mathbf{x}^{(k)}) - P^{(k)}(\mathbf{x}_*^{(k)}) \leq \alpha^{(k)} \quad \textbf{or} \quad \exists g_i^{(k)} \in \partial_{x_i} P^{(k)}(\mathbf{x})|_{\mathbf{x}=\mathbf{x}^{(k)}} \ \text{ s.t. } \max_{i\in\mathcal{N}} \|g_i^{(k)}\|_2 \leq \frac{\xi^{(k)}}{\sqrt{N}}\right\}. \tag{40}$$

Clearly, for all random sequences $\{\mathbf{x}^{(k)}\}_{k=1}^{N(\epsilon)}$ satisfying random event Δ, Corollary 1 implies that $\big|\sum_{i\in\mathcal{N}} F_i\left(x_i^{(N(\epsilon))}\right) - F^*\big| \leq \epsilon$ and $\max_{(i,j)\in\mathcal{E}}\left\{\|x_i^{(N(\epsilon))} - x_j^{(N(\epsilon))}\|_2\right\} \leq \epsilon$. Hence, we have

$$\mathbf{P}(\epsilon) \geq \mathbb{P}(\Delta) \geq \prod_{k=1}^{N(\epsilon)} \mathbb{P}\left(P^{(k)}\big(\mathbf{x}^{(k)}\big) - P^{(k)}\big(\mathbf{x}_*^{(k)}\big) \leq \alpha^{(k)}\right) \geq 1 - p.$$

In the rest, we bound the total number of **ARBCD** iterations required by *asynchronous* variant of **DFAL** to compute $\mathbf{x}^{(N(\epsilon))}$. Note that $(1-p)^{\frac{1}{N(\epsilon)}}$ is a concave function for $p \in (0,1)$, and we have $(1-p)^{\frac{1}{N(\epsilon)}} \leq 1 - \frac{p}{N(\epsilon)}$. Therefore, Lemma 8 and the discussion after Lemma 8 together imply that the number of **ARBCD** iterations, $N^{(k)}$, to compute $\mathbf{x}^{(k)}$ satisfying either (39) or (9)(b) is bounded above for $1 \leq k \leq N(\epsilon)$ as follows

$$N^{(k)} \leq 2N\sqrt{\frac{2C^{(k)}}{\alpha^{(k)}}}\log\left(\frac{N(\epsilon)}{p}\right) = 2N\left(\log\left(\frac{1}{p}\right) + \log\log_{\frac{1}{c}}\left(\frac{\bar{C}}{\epsilon}\right)\right)\sqrt{\frac{2C^{(k)}}{\alpha^{(k)}}}, \tag{41}$$

with $C^{(k)} = P^{(k)}\left(\mathbf{x}^{(k-1)}\right) - P^{(k)}\left(\mathbf{x}_*^{(k)}\right) + \sum_{i\in\mathcal{N}} \frac{\tilde{L}_i^{(k)}}{2}\|x_i^{(k-1)} - x_{*i}^{(k)}\|_2^2$.

Convexity of $\{\rho_i\}_{i\in\mathcal{N}}$, and Lemma 2 imply that

$$P^{(k)}(\mathbf{x}^{(k-1)}) - P^{(k)}(\mathbf{x}_*^{(k)}) \leq \left\langle \lambda^{(k)} s^{(k)} + \nabla f^{(k)}(\mathbf{x}_*^{(k)}),\ \mathbf{x}^{(k-1)} - \mathbf{x}_*^{(k)} \right\rangle + \sum_{i\in\mathcal{N}} \frac{L_i^{(k)}}{2}\|x_i^{(k-1)} - x_{*i}^{(k)}\|_2^2,$$

where $s^{(k)} \in \partial\lambda^{(k)}\bar{\rho}(\mathbf{x})|_{\mathbf{x}=\mathbf{x}^{(k-1)}}$, and $\bar{\rho}(\mathbf{x}) = \sum_{i\in\mathcal{N}} \rho_i(x_i)$. Note that optimality conditions imply that $-\nabla f^{(k)}(\mathbf{x}_*^{(k)}) \in \partial\lambda^{(k)}\bar{\rho}(\mathbf{x})|_{\mathbf{x}=\mathbf{x}_*^{(k)}}$. Assumption 1 implies that $\|\nabla_{x_i} f^{(k)}(\mathbf{x}_*^{(k)})\|_2 \leq \lambda^{(k)} B_i$ and $\|s_i^{(k)}\|_2 \leq \lambda^{(k)} B_i$ for all $i \in \mathcal{N}$. Hence, for some $\tilde{C} > 0$, we have $C^{(k)} \leq \sum_{i\in\mathcal{N}} \left(\frac{L_i^{(k)} + \tilde{L}_i^{(k)}}{2} + 2\lambda^{(k)} B_i\right)\|x_i^{(k-1)} - x_{*i}^{(k)}\|_2^2 \leq \tilde{C} B_x^2$ for all $k \geq 1$. Consequently, we can bound the total number of **ARBCD** iterations to compute $\mathbf{x}^{(N(\epsilon))}$ as follows:

$$\sum_{k=1}^{N(\epsilon)} N^{(k)} \leq 2NB_x\sqrt{\frac{2\tilde{C}}{\alpha^{(0)}}}\left(\log\left(\frac{1}{p}\right) + \log\log_{\frac{1}{c}}\left(\frac{\bar{C}}{\epsilon}\right)\right)\sum_{k=1}^{N(\epsilon)} c^{-k}.$$

Since $N(\epsilon) = \log_{\frac{1}{c}}(\bar{C}/\epsilon)$, and $\sum_{k=1}^{N(\epsilon)} c^{-k} = \frac{\left(\frac{1}{c}\right)^{N(\epsilon)}-1}{1-c} = \bar{C}\epsilon^{-1}/(1-c)$. Hence, we can conclude that $\sum_{k=1}^{N(\epsilon)} N^{(k)} = \mathcal{O}\left(\frac{1}{\epsilon}\left(\log\left(\frac{1}{p}\right) + \log\log\left(\frac{1}{\epsilon}\right)\right)\right)$