Appendix

A Lemmas for Convergence Analysis of DiSPA

We first introduce the lemma that characterizes the optimality condition of local subproblem $\mathcal{L}_k^{(t)}(\mathbf{x}, \mathbf{y}_{[k]})$.

Lemma 2. *Assume that each $\phi_i(\cdot)$ is $(1/\gamma)$-smooth and $g(\cdot)$ is λ-strongly convex, $R = \max\{\|a_1\|_2, ...\|a_n\|_2\}$. Let $(\hat{\mathbf{x}}_k^{(t)}, \hat{\mathbf{y}}_{[k]}^{(t)})$ be the optimal solution of $\mathcal{L}_k^{(t)}(\mathbf{x}, \mathbf{y}_{[k]})$, $k = 1, 2, ..., K$. Based on the strongly convexity, we have*

$$\mathcal{L}_k^{(t)}(\mathbf{x}, \hat{\mathbf{y}}_{[k]}^{(t)}) \geq \mathcal{L}_k^{(t)}(\hat{\mathbf{x}}_k^{(t)}, \hat{\mathbf{y}}_{[k]}^{(t)}) + \left(\frac{1}{2\tau} + \frac{\lambda}{2}\right) \|\mathbf{x} - \hat{\mathbf{x}}_k^{(t)}\|^2, \quad \forall \mathbf{x} \in \mathbb{R}^d$$

$$\mathcal{L}_k^{(t)}(\hat{\mathbf{x}}_k^{(t)}, \mathbf{y}_{[k]}) \leq \mathcal{L}_k^{(t)}(\hat{\mathbf{x}}_k^{(t)}, \hat{\mathbf{y}}_{[k]}^{(t)}) - \left(\frac{K}{2\sigma n} + \frac{\gamma}{2n}\right) \|\mathbf{y}_{[k]} - \hat{\mathbf{y}}_{[k]}^{(t)}\|^2, \quad \forall \mathbf{y}_{[k]} \in \mathbb{R}^n$$

Proof. Based on the definition of the saddle point, we can notice that $-\mathcal{L}_k^{(t)}(\hat{\mathbf{x}}^{(t)}, \mathbf{y}_{[k]})$ is a $\left(\frac{K}{\sigma n} + \frac{\gamma}{n}\right)$-strongly convex function and minimized by $\hat{\mathbf{y}}_{[k]}^{(t)}$, which implies

$$-\mathcal{L}_k^{(t)}(\hat{\mathbf{x}}^{(t)}, \mathbf{y}_{[k]}) \geq -\mathcal{L}_k^{(t)}(\hat{\mathbf{x}}^{(t)}, \hat{\mathbf{y}}_{[k]}^{(t)}) + \left(\frac{K}{2\sigma n} + \frac{\gamma}{2n}\right) \|\mathbf{y}_{[k]} - \hat{\mathbf{y}}_{[k]}^{(t)}\|^2, \quad \forall \mathbf{y}_{[k]} \in \mathbb{R}^n$$

Also notice that $\mathcal{L}_k^{(t)}(\mathbf{x}, \hat{\mathbf{y}}_{[k]}^{(t)})$ is a $\left(\frac{1}{\tau} + \lambda\right)$-strongly convex function minimized by $\hat{\mathbf{x}}_k^{(t)}$, which implies

$$\mathcal{L}_k^{(t)}(\mathbf{x}, \hat{\mathbf{y}}_{[k]}^{(t)}) \geq \mathcal{L}_k^{(t)}(\hat{\mathbf{x}}_k^{(t)}, \hat{\mathbf{y}}_{[k]}^{(t)}) + \left(\frac{1}{2\tau} + \frac{\lambda}{2}\right) \|\mathbf{x} - \hat{\mathbf{x}}_k^{(t)}\|^2, \quad \forall \mathbf{x} \in \mathbb{R}^d$$

□

Based on the optimality condition of $\mathcal{L}_k^{(t)}(\mathbf{x}, \mathbf{y}_{[k]})$ and the central update $\mathbf{x}^{(t)}$ on central worker, we can get the connection between local subproblems and central update on central worker.

Lemma 3 (Relationship between local optimal solution and central update)**.** *Assume that each $\phi_i(\cdot)$ is $(1/\gamma)$-smooth and $g(\cdot)$ is λ-strongly convex, $R = \max\{\|a_1\|_2, ...\|a_n\|_2\}$. Let $(\hat{\mathbf{x}}_k^{(t)}, \hat{\mathbf{y}}_{[k]}^{(t)})$ be the optimal solution of $\mathcal{L}_k^{(t)}(\mathbf{x}, \mathbf{y}_{[k]})$, $k = 1, 2, ..., K$. Let $\mathbf{x}^{(t)} = \arg\min_{\mathbf{x} \in \mathbb{R}^d} \mathcal{C}^{(t)}(\mathbf{x})$, it holds that*

$$\frac{(K-1)^2}{4\sigma n K} \|\mathbf{y}^{(t-1)} - \mathbf{y}^{(t)}\|_2^2 + \mathbf{\Lambda}' \geq \left(\frac{3}{4\tau} + \lambda\right) \sum_{k=1}^{K} \|\mathbf{x}^{(t)} - \hat{\mathbf{x}}_k^{(t)}\|_2^2 \tag{9}$$

where $\mathbf{\Lambda}' = \frac{1}{n} \sum_{k=1}^{K} (\hat{\mathbf{y}}_{[k]}^{(t)} - \mathbf{y}_{[k]}^{(t)})^\top A^\top (\mathbf{x}^{(t)} - \hat{\mathbf{x}}_k^{(t)})$ *and* $1/\tau = 4\sigma R^2$*.*

Proof. Based on lemma 2, we could get that for $k = 1, 2, ..., K$,

$$\frac{1}{n} \sum_{j \in \mathcal{P}_k} \hat{y}_j^{(t)} \langle a_j, \mathbf{x}^{(t)} - \hat{\mathbf{x}}_k^{(t)} \rangle + \frac{1}{n} \sum_{i \notin \mathcal{P}_k} y_i^{(t-1)} \langle a_i, \mathbf{x}^{(t)} - \hat{\mathbf{x}}_k^{(t)} \rangle + g(\mathbf{x}^{(t)}) - g(\hat{\mathbf{x}}_k^{(t)}) + \frac{\|\mathbf{x}^{(t)} - \mathbf{x}^{(t-1)}\|_2^2}{2\tau} \geq \frac{\|\hat{\mathbf{x}}_k^{(t)} - \mathbf{x}^{(t-1)}\|_2^2}{2\tau} + \left(\frac{1}{2\tau} + \frac{\lambda}{2}\right) \|\mathbf{x}^{(t)} - \hat{\mathbf{x}}_k^{(t)}\|_2^2$$

Since $\mathbf{x}^{(t)}$ minimizes the function $\mathcal{C}^{(t)}(\mathbf{x})$, which means that for $k = 1, 2, ..., K$, we have

$$\frac{1}{n} \sum_{i=1}^{n} y_i^{(t)} \langle a_i, \hat{\mathbf{x}}_k^{(t)} - \mathbf{x}^{(t)} \rangle + g(\hat{\mathbf{x}}_k^{(t)}) - g(\mathbf{x}^{(t)}) + \frac{\|\hat{\mathbf{x}}_k^{(t)} - \mathbf{x}^{(t-1)}\|_2^2}{2\tau} \geq \frac{\|\mathbf{x}^{(t)} - \mathbf{x}^{(t-1)}\|_2^2}{2\tau} + \left(\frac{1}{2\tau} + \frac{\lambda}{2}\right) \|\mathbf{x}^{(t)} - \hat{\mathbf{x}}_k^{(t)}\|_2^2$$

Sum up the above two inequalities, we can get

$$\frac{1}{n}\sum_{j\in\mathcal{P}_k}\hat{y}_j^{(t)}\langle a_j,\mathbf{x}^{(t)}-\hat{\mathbf{x}}_k^{(t)}\rangle+\frac{1}{n}\sum_{i\notin\mathcal{P}_k}y_i^{(t-1)}\langle a_i,\mathbf{x}^{(t)}-\hat{\mathbf{x}}_k^{(t)}\rangle+\frac{1}{n}\sum_{i=1}^{n}y_i^{(t)}\langle a_i,\hat{\mathbf{x}}_k^{(t)}-\mathbf{x}^{(t)}\rangle\geq\left(\frac{1}{\tau}+\lambda\right)\|\mathbf{x}^{(t)}-\hat{\mathbf{x}}_k^{(t)}\|_2^2$$

$$\frac{1}{n}\sum_{j\in\mathcal{P}_k}\left(\hat{y}_j^{(t)}-y_j^{(t)}\right)\langle a_j,\mathbf{x}^{(t)}-\hat{\mathbf{x}}_k^{(t)}\rangle+\frac{1}{n}\sum_{i\notin\mathcal{P}_k}\left(y_i^{(t-1)}-y_i^{(t)}\right)\langle a_i,\mathbf{x}^{(t)}-\hat{\mathbf{x}}_k^{(t)}\rangle\geq\left(\frac{1}{\tau}+\lambda\right)\|\mathbf{x}^{(t)}-\hat{\mathbf{x}}_k^{(t)}\|_2^2$$

$$\frac{1}{n}(\hat{\mathbf{y}}_{[k]}^{(t)}-\mathbf{y}_{[k]}^{(t)})^\top A^\top(\mathbf{x}^{(t)}-\hat{\mathbf{x}}_k^{(t)})+\frac{1}{n}\sum_{l\neq k}(\mathbf{y}_{[l]}^{(t-1)}-\mathbf{y}_{[l]}^{(t)})^\top A^\top(\mathbf{x}^{(t)}-\hat{\mathbf{x}}_k^{(t)})\geq\left(\frac{1}{\tau}+\lambda\right)\|\mathbf{x}^{(t)}-\hat{\mathbf{x}}_k^{(t)}\|_2^2$$

We need to upper bound the second term on the left-hand-side of the above inequality. Since $\|a_i\|_2\leq R$, and we assume that $1/\tau=4\sigma R^2$, then

$$\begin{aligned}\left|\frac{1}{n}\sum_{l\neq k}(\mathbf{y}_{[l]}^{(t-1)}-\mathbf{y}_{[l]}^{(t)})^\top A^\top(\mathbf{x}^{(t)}-\hat{\mathbf{x}}_k^{(t)})\right| &\leq \frac{\|\mathbf{x}^{(t)}-\hat{\mathbf{x}}_k^{(t)}\|_2^2}{4\tau}+\frac{\|\sum_{l\neq k}A(\mathbf{y}_{[l]}^{(t-1)}-\mathbf{y}_{[l]}^{(t)})\|_2^2}{n^2/\tau}\\ &\leq \frac{\|\mathbf{x}^{(t)}-\hat{\mathbf{x}}_k^{(t)}\|_2^2}{4\tau}+\frac{(\sum_{i\notin\mathcal{P}_k}|y_i^{(t-1)}-y_i^{(t)}|\cdot\|a_i\|)^2}{4\sigma R^2 n^2}\\ &\leq \frac{\|\mathbf{x}^{(t)}-\hat{\mathbf{x}}_k^{(t)}\|_2^2}{4\tau}+\frac{(K-1)}{4\sigma nK}\sum_{l\neq k}\|\mathbf{y}_{[l]}^{(t-1)}-\mathbf{y}_{[l]}^{(t)}\|_2^2\end{aligned}$$

Combine the above two inequalities,

$$\frac{(K-1)}{4\sigma nK}\sum_{l\neq k}\|\mathbf{y}_{[l]}^{(t-1)}-\mathbf{y}_{[l]}^{(t)}\|_2^2+\frac{1}{n}(\hat{\mathbf{y}}_{[k]}^{(t)}-\mathbf{y}_{[k]}^{(t)})^\top A^\top(\mathbf{x}^{(t)}-\hat{\mathbf{x}}_k^{(t)})\geq\left(\frac{3}{4\tau}+\lambda\right)\|\mathbf{x}^{(t)}-\hat{\mathbf{x}}_k^{(t)}\|_2^2$$

Sum up the above inequality over $k=1,2,...,K$, and denote $\mathbf{\Lambda}'=\frac{1}{n}\sum_{k=1}^{K}(\hat{\mathbf{y}}_{[k]}^{(t)}-\mathbf{y}_{[k]}^{(t)})^\top A^\top(\mathbf{x}^{(t)}-\hat{\mathbf{x}}_k^{(t)})$, we have

$$\frac{(K-1)^2}{4\sigma nK}\|\mathbf{y}^{(t-1)}-\mathbf{y}^{(t)}\|_2^2+\mathbf{\Lambda}'\geq\left(\frac{3}{4\tau}+\lambda\right)\sum_{k=1}^{K}\|\mathbf{x}^{(t)}-\hat{\mathbf{x}}_k^{(t)}\|_2^2$$

□

Lemma 3 shows that the distance between $\mathbf{x}^{(t)}$ and $\hat{\mathbf{x}}_k^{(t)}$ can be control by the update of $\mathbf{y}^{(t)}$ in each iteration.

B Proofs of Convergence for DiSPA and A-DiSPA

B.1 Proof of Lemma 1

Proof. We start from characterizing the relationship between $\mathbf{x}^{(t)}$ and $\mathbf{x}^\star$ after the t-update in Algorithm 1. According to the definition of $\mathbf{x}^{(t)}$, we have $\mathcal{C}^{(t)}(\mathbf{x}^\star)\geq\mathcal{C}^{(t)}(\mathbf{x}^{(t)})+\left(\frac{1}{2\tau}+\frac{\lambda}{2}\right)\|\mathbf{x}^{(t)}-\mathbf{x}^\star\|_2^2$, i.e.,

$$\frac{1}{n}\sum_{i=1}^{n}y_i^{(t)}\langle a_i,\mathbf{x}^\star-\mathbf{x}^{(t)}\rangle+g(\mathbf{x}^\star)-g(\mathbf{x}^{(t)})+\frac{\|\mathbf{x}^\star-\mathbf{x}^{(t-1)}\|_2^2}{2\tau}\geq\frac{\|\mathbf{x}^{(t)}-\mathbf{x}^{(t-1)}\|_2^2}{2\tau}+\left(\frac{1}{2\tau}+\frac{\lambda}{2}\right)\|\mathbf{x}^{(t)}-\hat{\mathbf{x}}^\star\|_2^2 \quad (10)$$

We also could derive the inequality characterizing the relation between $\hat{\mathbf{y}}^{(t)}=\sum_{k=1}^{K}\hat{\mathbf{y}}_{[k]}^{(t)}$ and $\mathbf{y}^\star$. For $k=1,2,...,K$, we have

$$\hat{\mathbf{y}}_{[k]}^{(t)}=\arg\max_{\mathbf{y}_{[k]}\in\mathbb{R}^n}\mathcal{L}_k^{(t)}(\hat{\mathbf{x}}_k^{(t)},\mathbf{y}_{[k]})$$

Since $\phi_i(\cdot)is(1/\gamma)$-smooth, we have $\phi_i^*(\cdot)$ is γ-strongly convex, therefore, for $k=1,2,...,K$, $\forall j\in\mathcal{P}_k$

$$(\hat{y}_j^{(t)}-y_j^\star)\langle a_j,\hat{\mathbf{x}}_k^{(t)}\rangle+\left(\phi_j^*(y_j^\star)-\phi_j^*(\hat{y}_j^{(t)})\right)+\frac{K(y_j^\star-y_j^{(t-1)})^2}{2\sigma}\geq\frac{K(\hat{y}_j^{(t)}-y_j^{(t-1)})^2}{2\sigma}+\left(\frac{K}{2\sigma}+\frac{\gamma}{2}\right)(\hat{y}_j^{(t)}-y_j^\star)^2$$

Sum up the above inequality over j, we have

$$\sum_{j\in\mathcal{P}_k}(\hat{y}_j^{(t)}-y_j^\star)\langle a_j,\hat{\mathbf{x}}_k^{(t)}\rangle+\sum_{j\in\mathcal{P}_k}\left(\phi_j^*(y_j^\star)-\phi_j^*(\hat{y}_j^{(t)})\right)+\frac{K\|\mathbf{y}_{[k]}^\star-\mathbf{y}_{[k]}^{(t-1)}\|_2^2}{2\sigma}\geq\frac{K\|\hat{\mathbf{y}}_{[k]}^{(t)}-\mathbf{y}_{[k]}^{(t-1)}\|_2^2}{2\sigma}+\left(\frac{K}{2\sigma}+\frac{\gamma}{2}\right)\|\mathbf{y}_{[k]}^\star-\hat{\mathbf{y}}_{[k]}^{(t)}\|_2^2$$

Sum up the above inequality over $k=1,2,...,K$ and multiplying both sides by $1/n$,

$$\begin{aligned}&\frac{1}{n}\sum_{k=1}^K\sum_{j\in\mathcal{P}_k}(\hat{y}_j^{(t)}-y_j^\star)\langle a_j,\hat{\mathbf{x}}_k^{(t)}\rangle+\frac{1}{n}\sum_{i=1}^n\left(\phi_i^*(y_j^\star)-\phi_i^*(\hat{y}_j^{(t)})\right)+\frac{K\|\mathbf{y}^\star-\mathbf{y}^{(t-1)}\|_2^2}{2\sigma n}\\&\geq\frac{K\|\hat{\mathbf{y}}^{(t)}-\mathbf{y}^{(t-1)}\|_2^2}{2\sigma n}+\left(\frac{K}{2\sigma n}+\frac{\gamma}{2n}\right)\|\mathbf{y}^\star-\hat{\mathbf{y}}^{(t)}\|_2^2\end{aligned}\tag{11}$$

In addition, we consider a combination of the saddle-point function values at different points, we have

$$\begin{aligned}&\left(f(\mathbf{x}^{(t)},\mathbf{y}^\star)-f(\mathbf{x}^\star,\mathbf{y}^\star)\right)+\left(f(\mathbf{x}^\star,\mathbf{y}^\star)-f(\mathbf{x}^\star,\mathbf{y}^{(t)})\right)\\&=\left(\frac{1}{n}\sum_{i=1}^n y_i^\star\langle a_i,\mathbf{x}^{(t)}\rangle-\frac{1}{n}\sum_{i=1}^n\phi_i^*(y_i^\star)+g(\mathbf{x}^{(t)})\right)-\left(\frac{1}{n}\sum_{i=1}^n y_i^{(t)}\langle a_i,\mathbf{x}^\star\rangle-\frac{1}{n}\sum_{i=1}^n\phi_i^*(y_i^{(t)})+g(\mathbf{x}^\star)\right)\\&=\left(g(\mathbf{x}^{(t)})-g(\mathbf{x}^\star)\right)+\frac{1}{n}\sum_{i=1}^n\left(\phi_i^*(y_i^{(t)})-\phi_i^*(y_i^\star)\right)+\left(\frac{1}{n}\sum_{i=1}^n y_i^\star\langle a_i,\mathbf{x}^{(t)}\rangle-\frac{1}{n}\sum_{i=1}^n y_i^{(t)}\langle a_i,\mathbf{x}^\star\rangle\right)\end{aligned}$$

Then we add (10) and (11) to the above equality, which implies

$$\begin{aligned}&\frac{1}{n}\sum_{k=1}^K\sum_{j\in\mathcal{P}_k}(\hat{y}_j^{(t)}-y_j^\star)\langle a_j,\hat{\mathbf{x}}_k\rangle+\frac{1}{n}\sum_{i=1}^n y_i^{(t)}\langle a_i,\mathbf{x}^\star-\mathbf{x}^{(t)}\rangle+\frac{1}{n}\sum_{i=1}^n y_i^\star\langle a_i,\mathbf{x}^{(t)}\rangle-\frac{1}{n}\sum_{i=1}^n y_i^{(t)}\langle a_i,\mathbf{x}^\star\rangle\\&\quad+\frac{1}{n}\sum_{i=1}^n\left(\phi_i^*(y_i^{(t)})-\phi_i^*(\hat{y}_i^{(t)})\right)+\frac{\|\mathbf{x}^\star-\mathbf{x}^{(t-1)}\|_2^2}{2\tau}+\frac{K\|\mathbf{y}^\star-\mathbf{y}^{(t-1)}\|_2^2}{2\sigma n}\\&\geq\left(f(\mathbf{x}^{(t)},\mathbf{y}^\star)-f(\mathbf{x}^\star,\mathbf{y}^\star)\right)+\left(f(\mathbf{x}^\star,\mathbf{y}^\star)-f(\mathbf{x}^\star,\mathbf{y}^{(t)})\right)\\&\quad+\frac{\|\mathbf{x}^{(t)}-\mathbf{x}^{(t-1)}\|_2^2}{2\tau}+\frac{K\|\hat{\mathbf{y}}^{(t)}-\mathbf{y}^{(t-1)}\|_2^2}{2\sigma n}+\left(\frac{1}{2\tau}+\frac{\lambda}{2}\right)\|\mathbf{x}^{(t)}-\mathbf{x}^\star\|_2^2+\left(\frac{K}{2\sigma n}+\frac{\gamma}{2n}\right)\|\mathbf{y}^\star-\hat{\mathbf{y}}^{(t)}\|_2^2\end{aligned}$$

which implies that

$$\begin{aligned}&\frac{1}{n}\sum_{k=1}^K\sum_{j\in\mathcal{P}_k}(\hat{y}_j^{(t)}-y_j^\star)\langle a_j,\hat{\mathbf{x}}_k-\mathbf{x}^{(t)}\rangle+\frac{1}{n}\sum_{i=1}^n(\hat{y}_i^{(t)}-y_i^{(t)})\langle a_i,\mathbf{x}^{(t)}\rangle+\frac{1}{n}\sum_{i=1}^n\left(\phi_i^*(y_i^{(t)})-\phi_i^*(\hat{y}_i)\right)\\&\quad+\frac{\|\mathbf{x}^\star-\mathbf{x}^{(t-1)}\|_2^2}{2\tau}\|^2+\frac{K\|\mathbf{y}^\star-\mathbf{y}^{(t-1)}\|_2^2}{2\sigma n}\\&\geq\left(f(\mathbf{x}^{(t)},\mathbf{y}^\star)-f(\mathbf{x}^\star,\mathbf{y}^\star)\right)+\left(f(\mathbf{x}^\star,\mathbf{y}^\star)-f(\mathbf{x}^\star,\mathbf{y}^{(t)})\right)+\frac{\|\mathbf{x}^{(t)}-\mathbf{x}^{(t-1)}\|_2^2}{2\tau}+\frac{K\|\hat{\mathbf{y}}^{(t)}-\mathbf{y}^{(t-1)}\|_2^2}{2\sigma n}\\&\quad+\left(\frac{1}{2\tau}+\frac{\lambda}{2}\right)\|\mathbf{x}^{(t)}-\mathbf{x}^\star\|_2^2+\left(\frac{K}{2\sigma n}+\frac{\gamma}{2n}\right)\|\mathbf{y}^\star-\hat{\mathbf{y}}^{(t)}\|_2^2\end{aligned}$$

We need to upper bound the first term on the left-hand-side of the above inequality, assume that $1/\tau = 4\sigma R^2$, we have

$$
\begin{aligned}
|\frac{1}{n}\sum_{j\in\mathcal{P}_k}(\hat{y}_j^{(t)} - y_j^\star)\langle a_j, \hat{\mathbf{x}}_k^{(t)} - \mathbf{x}^{(t)}\rangle| &= |\frac{1}{n}(\hat{\mathbf{y}}_{[k]}^{(t)} - \mathbf{y}_{[k]}^\star)^\top A^\top(\hat{\mathbf{x}}_k^{(t)} - \mathbf{x}^{(t)})| \\
&\le \frac{\|\hat{\mathbf{x}}_k^{(t)} - \mathbf{x}^{(t)}\|_2^2}{4\tau} + \frac{\|A(\hat{\mathbf{y}}_{[k]}^{(t)} - \mathbf{y}_{[k]}^\star)\|_2^2}{n^2/\tau} \\
&\le \frac{\|\hat{\mathbf{x}}_k^{(t)} - \mathbf{x}^{(t)}\|_2^2}{4\tau} + \frac{(\sum_{j\in\mathcal{P}_k}|\hat{y}_j^{(t)} - y_j^\star|\cdot\|a_j\|)^2}{4\sigma R^2 n^2} \\
&\le \frac{\|\hat{\mathbf{x}}_k^{(t)} - \mathbf{x}^{(t)}\|_2^2}{4\tau} + \frac{\|\hat{\mathbf{y}}_{[k]}^{(t)} - \mathbf{y}_{[k]}^\star\|_2^2}{4\sigma n K}
\end{aligned}
$$

then we can get the upper bound

$$
\frac{1}{n}\sum_{k=1}^K\sum_{j\in\mathcal{P}_k}(\hat{y}_j^{(t)} - y_j^\star)\langle a_j, \hat{\mathbf{x}}_k - \mathbf{x}^{(t)}\rangle \le \sum_{k=1}^K \frac{\|\hat{\mathbf{x}}_k^{(t)} - \mathbf{x}^{(t)}\|_2^2}{4\tau} + \frac{\|\hat{\mathbf{y}}^{(t)} - \mathbf{y}^\star\|_2^2}{4\sigma n K}
$$

next we denote that

$$
\mathbf{\Lambda}'' = \frac{1}{n}\sum_{i=1}^n(\hat{y}_i^{(t)} - y_i^{(t)})\langle a_i, \mathbf{x}^{(t)}\rangle + \frac{1}{n}\sum_{i=1}^n\left(\phi_i^*(y_i^{(t)}) - \phi_i^*(\hat{y}_i)\right)
$$

Combining the above inequality and equation, we derive that

$$
\begin{aligned}
&\sum_{k=1}^K\frac{\|\hat{\mathbf{x}}_k^{(t)} - \mathbf{x}^{(t)}\|_2^2}{4\tau} + \frac{\|\mathbf{x}^\star - \mathbf{x}^{(t-1)}\|_2^2}{2\tau} + \frac{K\|\mathbf{y}^\star - \mathbf{y}^{(t-1)}\|_2^2}{2\sigma n} + \mathbf{\Lambda}'' \\
\ge &\left(f(\mathbf{x}^{(t)}, \mathbf{y}^\star) - f(\mathbf{x}^\star, \mathbf{y}^\star)\right) + \left(f(\mathbf{x}^\star, \mathbf{y}^\star) - f(\mathbf{x}^\star, \mathbf{y}^{(t)})\right) + \frac{\|\mathbf{x}^{(t)} - \mathbf{x}^{(t-1)}\|_2^2}{2\tau} + \frac{K\|\hat{\mathbf{y}}^{(t)} - \mathbf{y}^{(t-1)}\|_2^2}{2\sigma n} \\
&+ \left(\frac{1}{2\tau} + \frac{\lambda}{2}\right)\|\mathbf{x}^{(t)} - \mathbf{x}^\star\|_2^2 + \left(\frac{K}{2\sigma n} + \frac{\gamma}{2n} - \frac{1}{4\sigma n K}\right)\|\mathbf{y}^\star - \hat{\mathbf{y}}^{(t)}\|_2^2
\end{aligned}
$$

Based on the inequality (9) in lemma 3, we could get that

$$
\begin{aligned}
&\sum_{k=1}^K\frac{\|\hat{\mathbf{x}}_k^{(t)} - \mathbf{x}^{(t)}\|_2^2}{4\tau} + \frac{(K-1)^2}{4\sigma n K}\|\mathbf{y}^{(t-1)} - \mathbf{y}^{(t)}\|_2^2 + \frac{\|\mathbf{x}^\star - \mathbf{x}^{(t-1)}\|_2^2}{2\tau} + \frac{K\|\mathbf{y}^\star - \mathbf{y}^{(t-1)}\|_2^2}{2\sigma n} + \mathbf{\Lambda}' + \mathbf{\Lambda}'' \\
\ge &\left(f(\mathbf{x}^{(t)}, \mathbf{y}^\star) - f(\mathbf{x}^\star, \mathbf{y}^\star)\right) + \left(f(\mathbf{x}^\star, \mathbf{y}^\star) - f(\mathbf{x}^\star, \mathbf{y}^{(t)})\right) + \frac{\|\mathbf{x}^{(t)} - \mathbf{x}^{(t-1)}\|_2^2}{2\tau} + \frac{K\|\hat{\mathbf{y}}^{(t)} - \mathbf{y}^{(t-1)}\|_2^2}{2\sigma n} \\
&+ \left(\frac{3}{4\tau} + \lambda\right)\sum_{k=1}^K\|\mathbf{x}^{(t)} - \hat{\mathbf{x}}_k^{(t)}\|_2^2 + \left(\frac{1}{2\tau} + \frac{\lambda}{2}\right)\|\mathbf{x}^{(t)} - \mathbf{x}^\star\|_2^2 + \left(\frac{K}{2\sigma n} + \frac{\gamma}{2n} - \frac{1}{4\sigma n K}\right)\|\mathbf{y}^\star - \hat{\mathbf{y}}^{(t)}\|_2^2
\end{aligned}
$$

Since we can rewrite $\|\mathbf{y} - \hat{\mathbf{y}}^{(t)}\|_2^2 = \|\mathbf{y} - \mathbf{y}^{(t)}\|_2^2 + \|\mathbf{y}^{(t)} - \hat{\mathbf{y}}^{(t)}\|_2^2 + 2(\mathbf{y} - \mathbf{y}^{(t)})^\top(\mathbf{y}^{(t)} - \hat{\mathbf{y}}^{(t)})$, then we denote that

$$
\begin{aligned}
\mathbf{\Lambda}''' = &-\left(\frac{K}{2\sigma n} + \frac{\gamma}{2n} - \frac{1}{4\sigma n K}\right)\left(\|\mathbf{y}^{(t)} - \hat{\mathbf{y}}^{(t)}\|_2^2 + 2(\mathbf{y}^\star - \mathbf{y}^{(t)})^\top(\mathbf{y}^{(t)} - \hat{\mathbf{y}}^{(t)})\right) \\
&- \frac{K}{2\sigma n}\left(\|\mathbf{y}^{(t)} - \hat{\mathbf{y}}^{(t)}\|_2^2 + 2(\mathbf{y}^{(t-1)} - \mathbf{y}^{(t)})^\top(\mathbf{y}^{(t)} - \hat{\mathbf{y}}^{(t)})\right)
\end{aligned}
$$

Based on the definition of $\mathbf{\Lambda}'''$, we derive that

$$
\begin{aligned}
&\frac{\|\mathbf{x}^\star - \mathbf{x}^{(t-1)}\|_2^2}{2\tau} + \frac{K\|\mathbf{y}^\star - \mathbf{y}^{(t-1)}\|_2^2}{2\sigma n} + \mathbf{\Lambda}' + \mathbf{\Lambda}'' + \mathbf{\Lambda}''' \\
\ge &\left(f(\mathbf{x}^{(t)}, \mathbf{y}^\star) - f(\mathbf{x}^\star, \mathbf{y}^\star)\right) + \left(f(\mathbf{x}^\star, \mathbf{y}^\star) - f(\mathbf{x}^\star, \mathbf{y}^{(t)})\right) \\
&+ \left(\frac{1}{2\tau} + \frac{\lambda}{2}\right)\|\mathbf{x}^{(t)} - \mathbf{x}^\star\|_2^2 + \left(\frac{K}{2\sigma n} + \frac{\gamma}{2n} - \frac{1}{4\sigma n K}\right)\|\mathbf{y}^\star - \mathbf{y}^{(t)}\|_2^2 + \mathbf{\Lambda}^\sharp
\end{aligned}
$$

where we define

$$\mathbf{\Lambda}^\sharp = \frac{\|\mathbf{x}^{(t)} - \mathbf{x}^{(t-1)}\|_2^2}{2\tau} + \left(\frac{1}{2\tau} + \lambda\right)\sum_{k=1}^{K}\|\mathbf{x}^{(t)} - \hat{\mathbf{x}}_k^{(t)}\|_2^2 + \left(\frac{K}{2\sigma n} - \frac{(K-1)^2}{4\sigma n K}\right)\|\mathbf{y}^{(t)} - \mathbf{y}^{(t-1)}\|_2^2$$

Assume that $\Theta \geq 0$, then we can obtain

$$\begin{aligned}
&\Theta\left(f(\mathbf{x}^{(t-1)}, \mathbf{y}^\star) - f(\mathbf{x}^\star, \mathbf{y}^\star)\right) + \Theta\left(f(\mathbf{x}^\star, \mathbf{y}^\star) - f(\mathbf{x}^\star, \mathbf{y}^{(t-1)})\right)\\
&\frac{\|\mathbf{x}^\star - \mathbf{x}^{(t-1)}\|_2^2}{2\tau} + \frac{K\|\mathbf{y}^\star - \mathbf{y}^{(t-1)}\|_2^2}{2\sigma n} + \mathbf{\Lambda}\\
\geq &\left(f(\mathbf{x}^{(t)}, \mathbf{y}^\star) - f(\mathbf{x}^\star, \mathbf{y}^\star)\right) + \left(f(\mathbf{x}^\star, \mathbf{y}^\star) - f(\mathbf{x}^\star, \mathbf{y}^{(t)})\right) + \Theta\left(f(\mathbf{x}^{(t-1)}, \mathbf{y}^\star) - f(\mathbf{x}^\star, \mathbf{y}^{(t-1)})\right)\\
&+ \left(\frac{1}{2\tau} + \frac{\lambda}{2}\right)\|\mathbf{x}^{(t)} - \mathbf{x}^\star\|_2^2 + \left(\frac{K}{2\sigma n} + \frac{\gamma}{2n} - \frac{1}{4\sigma n K}\right)\|\mathbf{y}^\star - \mathbf{y}^{(t)}\|_2^2 + \mathbf{\Lambda}^\sharp
\end{aligned}$$

where $\mathbf{\Lambda} = \mathbf{\Lambda}' + \mathbf{\Lambda}'' + \mathbf{\Lambda}'''$ and $\mathbf{\Lambda}^\sharp > 0$.

In order to get the convergence guarantee of our algorithm, we need to get the upper bound of $\mathbf{\Lambda}$, based on the definition of $\mathbf{\Lambda}$, we could get

$$\begin{aligned}
\mathbf{\Lambda} \leq &\frac{1}{n}\sum_{k=1}^{K}(\hat{\mathbf{y}}_{[k]}^{(t)} - \mathbf{y}_{[k]}^{(t)})^\top A^\top(\mathbf{x}^{(t)} - \hat{\mathbf{x}}_k^{(t)}) + \frac{1}{n}(\hat{\mathbf{y}}^{(t)} - \mathbf{y}^{(t)})^\top A^\top \mathbf{x}^{(t)} + \frac{1}{n}\sum_{i=1}^{n}\left(\phi_i^*(y_i^{(t)}) - \phi_i^*(\hat{y}_i)\right)\\
&+ \left(\frac{K}{\sigma n} + \frac{\gamma}{n} - \frac{1}{2\sigma n K}\right)(\mathbf{y}^\star - \mathbf{y}^{(t)})^\top(\hat{\mathbf{y}}^{(t)} - \mathbf{y}^{(t)}) + \frac{K}{\sigma n}(\mathbf{y}^{(t-1)} - \mathbf{y}^{(t)})^\top(\hat{\mathbf{y}}^{(t)} - \mathbf{y}^{(t)})
\end{aligned}$$

based on the assumption, then we could get that

$$\mathbf{\Lambda} \leq \frac{M}{n}\|\hat{\mathbf{y}}^{(t)} - \mathbf{y}^{(t)}\|_2 + \sum_{k=1}^{K}\left(f(\hat{\mathbf{x}}_k^{(t)}, \hat{\mathbf{y}}_{[k]}^{(t)}) - f(\hat{\mathbf{x}}_k^{(t)}, \mathbf{y}_{[k]}^{(t)})\right)$$

where $M = 4\sqrt{n}R\,\Omega_{\mathbf{x}} + \left(\frac{3K}{\sigma} + \gamma\right)\Omega_{\mathbf{y}} = c'\Omega_{\mathbf{x}} + c''\Omega_{\mathbf{y}}$ as defined in Assumption 5. Based on the assumption

$$\mathbb{E}[\mathbf{\Lambda}] \leq \hat{\Theta}\frac{M}{n}\|\mathbf{y}^{(t-1)} - \hat{\mathbf{y}}^{(t)}\|_2 + \tilde{\Theta}\sum_{k=1}^{K}\left(f(\hat{\mathbf{x}}_k^{(t)}, \hat{\mathbf{y}}_{[k]}^{(t)}) - f(\hat{\mathbf{x}}_k^{(t)}, \mathbf{y}_{[k]}^{(t-1)})\right)$$

where $\hat{\Theta} < 1$, $\tilde{\Theta} < 1$ are defined in Assumption 4, and define the parameter Θ as

$$\Theta = \max\left\{\frac{1}{1 + \left(\frac{\sigma\gamma}{K} - \frac{1}{2K^2}\right)}, \frac{1}{1 + \tau\lambda}\right\} \tag{12}$$

based on Assumption 5, and we could get that

$$\mathbb{E}[\blacktriangle] = \mathbb{E}[\frac{M}{n}\|\hat{\mathbf{y}}^{(t)} - \mathbf{y}^{(t)}\|_2 + \sum_{k=1}^{K}\left(f(\hat{\mathbf{x}}_k^{(t)}, \hat{\mathbf{y}}_{[k]}^{(t)}) - f(\hat{\mathbf{x}}_k^{(t)}, \mathbf{y}_{[k]}^{(t)})\right)] \leq \Theta\left(f(\mathbf{x}^{(t-1)}, \mathbf{y}^\star) - f(\mathbf{x}^\star, \mathbf{y}^{(t-1)})\right)$$

Then we could get our finial conclusion

$$\mathbb{E}[\Delta^{(t)}] \leq \Theta\mathbb{E}[\Delta^{(t-1)}]$$

where we require that

$$\sigma > \frac{1}{2K\gamma}, \quad \tau\sigma = \frac{1}{4R^2} \tag{13}$$

□

B.2 Proof of Theorem 1

Proof. By Lemma 1, for each $t > 0$, we have $\mathbb{E}[\Delta^{(t)}] \leq \Theta^t \mathbb{E}[\Delta^{(0)}]$, according to the definition of C, we have

$$\mathbb{E}[\|\mathbf{x}^{(t)} - \mathbf{x}^\star\|_2^2] \leq \Theta^t C \tag{14}$$

In order to obtain $\Theta^T C \leq \epsilon$, which needs

$$T \geq -\frac{\log{(C/\epsilon)}}{\log{(\Theta)}}$$

Suppose the parameters τ, σ are set as

$$\sigma = \frac{1}{\gamma}, \quad \tau = \frac{\gamma}{4R^2}$$

Then according to the definition of Θ

$$\Theta = \max\left\{\frac{1}{1+\left(\frac{\sigma\gamma}{K} - \frac{1}{2K^2}\right)}, \frac{1}{1+\tau\lambda}\right\} = \max\left\{\frac{1}{1+\left(\frac{1}{K} - \frac{1}{2K^2}\right)}, \frac{1}{1+(\lambda\gamma)/(4R^2)}\right\}$$

As long as $\frac{1}{K} - \frac{1}{2K^2} \geq \frac{1}{K} - \frac{1}{2K} = \frac{1}{2K} \geq \frac{\lambda\gamma}{4R^2}$, since we assume $K \leq \kappa$. Then we have

$$\Theta = \frac{1}{1+(\lambda\gamma)/(4R^2)} = 1 - \frac{1}{1+(4R^2)/(\lambda\gamma)}$$

Thus we can get

$$T \geq -\frac{\log{(C/\epsilon)}}{\log{(\Theta)}} = \frac{\log{(C/\epsilon)}}{-\log\left(1 - \frac{1}{1+(4R^2)/(\lambda\gamma)}\right)}$$

where we apply the inequality $\log(1-x) \leq -x$ in the last inequality and we could get the conclusion. $\square$

B.3 Convergence guarantee of primal-dual gap

Next we derive the convergence rate of primal-dual gap based on Theorem 1.

Lemma 4 (Yu et al. (2015))**.** *Suppose Assumption 1,2 holds, Let $(\mathbf{x}^\star, \mathbf{y}^\star)$ is the unique saddle-point of $f(\mathbf{x}, \mathbf{y})$, and $R = \max_{1\leq i\leq n} \|a_i\|_2$. Then for any point $(\mathbf{x}, \mathbf{y}) \in dom(g) \times dom(\phi^*)$, we have*

$$\mathcal{P}(\mathbf{x}) \leq f(\mathbf{x}, \mathbf{y}^\star) + \frac{R^2}{2\gamma}\|\mathbf{x} - \mathbf{x}^\star\|_2^2, \ \mathcal{D}(\mathbf{y}) \geq f(\mathbf{x}^\star, \mathbf{y}) - \frac{R^2}{2\lambda n}\|\mathbf{y} - \mathbf{y}^\star\|_2^2 \tag{15}$$

Corollary 2. *Suppose Assumption A holds and the parameters τ, σ, Θ are set as in (12), (13). Then the iterates of Algorithm 1 satisfy*

$$\mathbb{E}[\mathcal{P}(\mathbf{x}^{(T)}) - \mathcal{D}(\mathbf{y}^{(T)})] \leq \epsilon$$

it suffices to have the number of communication iterations T satisfy,

$$T \geq \left(1 + \frac{4R^2}{\lambda\gamma}\right) \log\left(\left(1 + \frac{R^2}{\lambda\gamma}\right)\frac{C}{\epsilon}\right)$$

.

Proof. Based on Lemma 4, we have

$$\begin{aligned}
\mathcal{P}(\mathbf{x}^{(t)}) - \mathcal{D}(\mathbf{y}^{(t)}) &= \mathcal{P}(\mathbf{x}^{(t)}) - \mathcal{P}(\mathbf{x}^\star) + \mathcal{D}(\mathbf{y}^\star) - \mathcal{D}(\mathbf{y}^{(t)}) \\
&\leq f(\mathbf{x}^{(t)}, \mathbf{y}^\star) + \frac{R^2}{2\gamma}\|\mathbf{x}^{(t)} - \mathbf{x}^\star\|_2^2 - f(\mathbf{x}^\star, \mathbf{y}^{(t)}) + \frac{R^2}{2\lambda n}\|\mathbf{y}^{(t)} - \mathbf{y}^\star\|_2^2 \\
&\leq \left(1 + \frac{R^2}{\lambda\gamma}\right)\Delta^{(t)}
\end{aligned}$$

Similar to Theorem 1, we could get the conclusion.

$\square$

B.4 Proof of Theorem 2

Proof. If the parameters τ, σ in Algorithm 2 are set as

$$\tau = \frac{\gamma}{4R^2}, \quad \sigma = \frac{1}{\gamma}$$

then based on Lemma 1, we have

$$\Theta = \max\left\{\frac{1}{1+(\frac{\sigma\gamma}{K}-\frac{1}{2K^2})}, \frac{1}{1+\tau\lambda}\right\} = \max\left\{\frac{1}{1+(\frac{1}{K}-\frac{1}{2K^2})}, \frac{1}{1+\frac{\lambda\gamma}{4R^2}}\right\}$$

assume that $K \geq 2$, and $(\frac{1}{K}-\frac{1}{2K^2}) \geq \frac{\lambda\gamma}{4R^2}$, which means we focus on the situation that $\kappa = \frac{R^2}{\lambda\gamma} \gg 1$. Then we could get that

$$\Theta = \frac{1}{1+\frac{\lambda\gamma}{4R^2}} = 1 - \frac{1}{1+\frac{4R^2}{\lambda\gamma}}$$

Based on Corollary 2, we could get that

$$\mathcal{P}(\mathbf{x}^{(t)}) - \mathcal{P}(\mathbf{x}^\star) \leq \mathcal{P}(\mathbf{x}^{(t)}) - \mathcal{D}(\mathbf{y}^{(t)}) \leq \left(1+\frac{R^2}{\lambda\gamma}\right)\Delta^{(t)} \leq C'\Theta^t\left(\mathcal{P}(\mathbf{x}^{(0)}) - \mathcal{P}(\mathbf{x}^\star)\right)$$

where C' is a constant and $C' > 0$ and $\mathcal{P}(\cdot)$ is defined in (1). If we apply DiSPA on $f^{(t)}$ defined in Section 5.1, then based on Proposition 3.2 of Lin et al. (2015), the $\tau_{\mathcal{M}}$ defined in Proposition 3.2 equals

$$\tau_{\mathcal{M}} = \frac{1}{1+\frac{4R^2}{(\lambda+\vartheta)\gamma}}$$

where ϑ is the parameter defined in Section 5.1. We could get that $\tau_{\mathcal{M}} \approx \frac{(\lambda+\vartheta)\gamma}{4R^2}$ since $\kappa \gg 1$. Now apply Proposition 3.2 in Lin et al. (2015), we could get the global linear rate of convergence withe parameter $\tau_{\mathcal{A},F}$,

$$\tau_{\mathcal{A},F} = \tilde{O}\left(\tau_{\mathcal{M}}\sqrt{\lambda}/\sqrt{\lambda+\vartheta}\right) = \tilde{O}\left(\frac{(\lambda+\vartheta)\,\gamma\sqrt{\lambda}}{4R^2\sqrt{\lambda+\vartheta}}\right)$$

if we take ϑ as $\vartheta = \frac{4R^2}{\gamma} - \lambda$, then we can get that the communication complexity of the A-DiSPA for achieving $\mathcal{P}(\mathbf{x}^{(T)}) - \mathcal{P}(\mathbf{x}^\star) < \epsilon$ is $\tilde{O}\left(\sqrt{\kappa}\log\left(\frac{1}{\epsilon}\right)\right)$. □