7. Appendix

7.1. Proof of Lemma 1

Proof. Invoking the optimality condition for (6), we have

$$\langle \mathbf{g}(\mathbf{x}^*) + s\nabla D(\mathbf{x}^*, \mathbf{u}), \mathbf{x} - \mathbf{x}^* \rangle \geq 0, \ \forall \mathbf{x} \in \mathcal{X},$$

which is equivalent to

$$\begin{aligned}\langle \mathbf{g}(\mathbf{x}^*), \mathbf{x}^* - \mathbf{x} \rangle &\leq s \langle \nabla D(\mathbf{x}^*, \mathbf{u}), \mathbf{x} - \mathbf{x}^* \rangle \\ &= s \langle \nabla \omega(\mathbf{x}^*) - \nabla \omega(\mathbf{u}), \mathbf{x} - \mathbf{x}^* \rangle \\ &= s \left[D(\mathbf{x}, \mathbf{u}) - D(\mathbf{x}, \mathbf{x}^*) - D(\mathbf{x}^*, \mathbf{u}) \right].\end{aligned}$$

□

7.2. Proof of Lemma 2

Proof. Due to the convexity of θ_1 and using the definition of $\boldsymbol{\delta}_k$, we have

$$\theta_1(\mathbf{x}_k) - \theta_1(\mathbf{x}) \leq \left\langle \theta_1'(\mathbf{x}_k), \mathbf{x}_k - \mathbf{x} \right\rangle = \left\langle \theta_1'(\mathbf{x}_k, \boldsymbol{\xi}_{k+1}), \mathbf{x}_{k+1} - \mathbf{x} \right\rangle + \left\langle \boldsymbol{\delta}_{k+1}, \mathbf{x} - \mathbf{x}_k \right\rangle + \left\langle \theta_1'(\mathbf{x}_k, \boldsymbol{\xi}_{k+1}), \mathbf{x}_k - \mathbf{x}_{k+1} \right\rangle. \quad (23)$$

Applying Lemma 1 to Line 1 of Alg.2 and taking $D(\mathbf{u}, \mathbf{v}) = \frac{1}{2}\|\mathbf{v} - \mathbf{u}\|^2$, we have

$$\begin{aligned}&\left\langle \theta_1'(\mathbf{x}_k, \boldsymbol{\xi}_{k+1}) + A^T \left[\beta(A\mathbf{x}_{k+1} + B\mathbf{y}_k - \mathbf{b}) - \boldsymbol{\lambda}_k \right], \mathbf{x}_{k+1} - \mathbf{x} \right\rangle \\ &\leq \frac{1}{2\eta_{k+1}} \left(\|\mathbf{x}_k - \mathbf{x}\|^2 - \|\mathbf{x}_{k+1} - \mathbf{x}\|^2 - \|\mathbf{x}_k - \mathbf{x}_{k+1}\|^2 \right)\end{aligned} \quad (24)$$

Combining (23) and (24) we have

$$\begin{aligned}&\theta_1(\mathbf{x}_k) - \theta_1(\mathbf{x}) + \left\langle \mathbf{x}_{k+1} - \mathbf{x}, -A^T \boldsymbol{\lambda}_{k+1} \right\rangle \\ &\overset{(23)}{\leq} \left\langle \theta_1'(\mathbf{x}_k, \boldsymbol{\xi}_{k+1}), \mathbf{x}_{k+1} - \mathbf{x} \right\rangle + \left\langle \boldsymbol{\delta}_{k+1}, \mathbf{x} - \mathbf{x}_k \right\rangle + \left\langle \theta_1'(\mathbf{x}_k, \boldsymbol{\xi}_{k+1}), \mathbf{x}_k - \mathbf{x}_{k+1} \right\rangle + \\ &\quad \left\langle \mathbf{x}_{k+1} - \mathbf{x}, A^T \left[\beta(A\mathbf{x}_{k+1} + B\mathbf{y}_{k+1} - \mathbf{b}) - \boldsymbol{\lambda}_k \right] \right\rangle \\ &= \left\langle \theta_1'(\mathbf{x}_k, \boldsymbol{\xi}_{k+1}) + A^T \left[\beta(A\mathbf{x}_{k+1} + B\mathbf{y}_k - \mathbf{b}) - \boldsymbol{\lambda}_k \right], \mathbf{x}_{k+1} - \mathbf{x} \right\rangle + \\ &\quad \left\langle \boldsymbol{\delta}_{k+1}, \mathbf{x} - \mathbf{x}_k \right\rangle + \left\langle \mathbf{x} - \mathbf{x}_{k+1}, \beta A^T B(\mathbf{y}_k - \mathbf{y}_{k+1}) \right\rangle + \left\langle \theta_1'(\mathbf{x}_k, \boldsymbol{\xi}_{k+1}), \mathbf{x}_k - \mathbf{x}_{k+1} \right\rangle \\ &\overset{(24)}{\leq} \frac{1}{2\eta_{k+1}} \left(\|\mathbf{x}_k - \mathbf{x}\|^2 - \|\mathbf{x}_{k+1} - \mathbf{x}\|^2 - \|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2 \right) + \left\langle \boldsymbol{\delta}_{k+1}, \mathbf{x} - \mathbf{x}_k \right\rangle + \\ &\quad \left\langle \mathbf{x} - \mathbf{x}_{k+1}, \beta A^T B(\mathbf{y}_k - \mathbf{y}_{k+1}) \right\rangle + \left\langle \theta_1'(\mathbf{x}_k, \boldsymbol{\xi}_{k+1}), \mathbf{x}_k - \mathbf{x}_{k+1} \right\rangle\end{aligned} \quad (25)$$

We handle the last two terms separately:

$$\begin{aligned}&\left\langle \mathbf{x} - \mathbf{x}_{k+1}, \beta A^T B(\mathbf{y}_k - \mathbf{y}_{k+1}) \right\rangle = \beta \left\langle A\mathbf{x} - A\mathbf{x}_{k+1}, B\mathbf{y}_k - B\mathbf{y}_{k+1} \right\rangle \\ &= \frac{\beta}{2} \left[\left(\|A\mathbf{x} + B\mathbf{y}_k - \mathbf{b}\|^2 - \|A\mathbf{x} + B\mathbf{y}_{k+1} - \mathbf{b}\|^2 \right) + \left(\|A\mathbf{x}_{k+1} + B\mathbf{y}_{k+1} - \mathbf{b}\|^2 - \|A\mathbf{x}_{k+1} + B\mathbf{y}_k - \mathbf{b}\|^2 \right) \right] \\ &\leq \frac{\beta}{2} \left(\|A\mathbf{x} + B\mathbf{y}_k - \mathbf{b}\|^2 - \|A\mathbf{x} + B\mathbf{y}_{k+1} - \mathbf{b}\|^2 \right) + \frac{1}{2\beta} \|\boldsymbol{\lambda}_{k+1} - \boldsymbol{\lambda}_k\|^2\end{aligned} \quad (26)$$

and

$$\left\langle \theta_1'(\mathbf{x}_k, \boldsymbol{\xi}_{k+1}), \mathbf{x}_k - \mathbf{x}_{k+1} \right\rangle \leq \frac{\eta_{k+1} \|\theta_1'(\mathbf{x}_k, \boldsymbol{\xi}_{k+1})\|^2}{2} + \frac{\|\mathbf{x}_k - \mathbf{x}_{k+1}\|^2}{2\eta_{k+1}}, \quad (27)$$

where the last step is due to Young's inequality. Inserting (26) and (27) into (25), we have

$$\begin{aligned}&\theta_1(\mathbf{x}_k) - \theta_1(\mathbf{x}) + \left\langle \mathbf{x}_{k+1} - \mathbf{x}, -A^T \boldsymbol{\lambda}_{k+1} \right\rangle \\ &\leq \frac{1}{2\eta_{k+1}} \left(\|\mathbf{x}_k - \mathbf{x}\|^2 - \|\mathbf{x}_{k+1} - \mathbf{x}\|^2 \right) + \frac{\eta_{k+1} \|\theta_1'(\mathbf{x}_k, \boldsymbol{\xi}_{k+1})\|^2}{2} + \left\langle \boldsymbol{\delta}_{k+1}, \mathbf{x} - \mathbf{x}_k \right\rangle \\ &+ \frac{\beta}{2} \left(\|A\mathbf{x} + B\mathbf{y}_k - \mathbf{b}\|^2 - \|A\mathbf{x} + B\mathbf{y}_{k+1} - \mathbf{b}\|^2 \right) + \frac{1}{2\beta} \|\boldsymbol{\lambda}_{k+1} - \boldsymbol{\lambda}_k\|^2,\end{aligned} \quad (28)$$

Due to the optimality condition of Line 2 in Alg.2 and the convexity of θ_2, we have

$$\theta_2(\mathbf{y}_{k+1}) - \theta_2(\mathbf{y}) + \left\langle \mathbf{y}_{k+1} - \mathbf{y}, -B^T \boldsymbol{\lambda}_{k+1} \right\rangle \leq 0. \tag{29}$$

Using Line 3 in Alg.2, we have

$$\begin{aligned}
&\langle \boldsymbol{\lambda}_{k+1} - \boldsymbol{\lambda}, A\mathbf{x}_{k+1} + B\mathbf{y}_{k+1} - \mathbf{b} \rangle \\
&= \frac{1}{\beta} \langle \boldsymbol{\lambda}_{k+1} - \boldsymbol{\lambda}, \boldsymbol{\lambda}_k - \boldsymbol{\lambda}_{k+1} \rangle \\
&= \frac{1}{2\beta} \left(\|\boldsymbol{\lambda} - \boldsymbol{\lambda}_k\|^2 - \|\boldsymbol{\lambda} - \boldsymbol{\lambda}_{k+1}\|^2 - \|\boldsymbol{\lambda}_{k+1} - \boldsymbol{\lambda}_k\|^2 \right)
\end{aligned} \tag{30}$$

Taking the summation of inequalities (28) (29) and (30), we obtain the result as desired. □

7.3. Proof of Theorem 1

Proof. (i). Invoking convexity of $\theta_1(\cdot)$ and $\theta_2(\cdot)$ and the monotonicity of operator $F(\cdot)$, we have $\forall \mathbf{w} \in \mathcal{W}$:

$$\begin{aligned}
\theta(\bar{\mathbf{u}}_t) - \theta(\mathbf{u}) + (\bar{\mathbf{w}}_t - \mathbf{w})^T F(\bar{\mathbf{w}}_t) &\leq \frac{1}{t} \sum_{k=1}^{t} \left[\theta_1(\mathbf{x}_{k-1}) + \theta_2(\mathbf{y}_k) - \theta(\mathbf{u}) + (\mathbf{w}_k - \mathbf{w})^T F(\mathbf{w}_k) \right] \\
&= \frac{1}{t} \sum_{k=0}^{t-1} \left[\theta_1(\mathbf{x}_k) + \theta_2(\mathbf{y}_{k+1}) - \theta(\mathbf{u}) + (\mathbf{w}_{k+1} - \mathbf{w})^T F(\mathbf{w}_{k+1}) \right]
\end{aligned} \tag{31}$$

Applying Lemma 2 at the optimal solution $(\mathbf{x}, \mathbf{y}) = (\mathbf{x}_*, \mathbf{y}_*)$, we can derive from (31) that, $\forall \boldsymbol{\lambda}$

$$\begin{aligned}
&\theta(\bar{\mathbf{u}}_t) - \theta(\mathbf{u}_*) + (\bar{\mathbf{x}}_t - \mathbf{x}_*)^T (-A^T \bar{\boldsymbol{\lambda}}_t) + (\bar{\mathbf{y}}_t - \mathbf{y}_*)^T (-B^T \bar{\boldsymbol{\lambda}}_t) + (\bar{\boldsymbol{\lambda}}_t - \boldsymbol{\lambda})^T (A\bar{\mathbf{x}}_t + B\bar{\mathbf{y}}_t - \mathbf{b}) \\
&\overset{(7)}{\leq} \frac{1}{t} \sum_{k=0}^{t-1} \left[\frac{\eta_{k+1} \|\theta_1'(\mathbf{x}_k, \boldsymbol{\xi}_{k+1})\|^2}{2} + \frac{1}{2\eta_{k+1}} \left(\|\mathbf{x}_k - \mathbf{x}_*\|^2 - \|\mathbf{x}_{k+1} - \mathbf{x}_*\|^2 \right) + \langle \boldsymbol{\delta}_{k+1}, \mathbf{x}_* - \mathbf{x}_k \rangle \right] \\
&\quad + \frac{1}{t} \left(\frac{\beta}{2} \|A\mathbf{x}_* + B\mathbf{y}_0 - \mathbf{b}\|^2 + \frac{1}{2\beta} \|\boldsymbol{\lambda} - \boldsymbol{\lambda}_0\|^2 \right) \\
&\leq \frac{1}{t} \sum_{k=0}^{t-1} \left[\frac{\eta_{k+1} \|\theta_1'(\mathbf{x}_k, \boldsymbol{\xi}_{k+1})\|^2}{2} + \langle \boldsymbol{\delta}_{k+1}, \mathbf{x}_* - \mathbf{x}_k \rangle \right] + \frac{1}{t} \left(\frac{D_{\mathcal{X}}^2}{2\eta_t} + \frac{\beta}{2} D_{\mathbf{y}_*, B}^2 + \frac{1}{2\beta} \|\boldsymbol{\lambda} - \boldsymbol{\lambda}_0\|_2^2 \right)
\end{aligned} \tag{32}$$

The above inequality is true for all $\boldsymbol{\lambda} \in \mathbb{R}^m$, hence it also holds in the ball $\mathcal{B}_0 = \{\boldsymbol{\lambda} : \|\boldsymbol{\lambda}\|_2 \leq \rho\}$. Combing with the fact that the optimal solution must also be feasible, it follows that

$$\begin{aligned}
&\max_{\boldsymbol{\lambda} \in \mathcal{B}_0} \left\{ \theta(\bar{\mathbf{u}}_t) - \theta(\mathbf{u}_*) + (\bar{\mathbf{x}}_t - \mathbf{x}_*)^T (-A^T \bar{\boldsymbol{\lambda}}_t) + (\bar{\mathbf{y}}_t - \mathbf{y}_*)^T (-B^T \bar{\boldsymbol{\lambda}}_t) + (\bar{\boldsymbol{\lambda}}_t - \boldsymbol{\lambda})^T (A\bar{\mathbf{x}}_t + B\bar{\mathbf{y}}_t - \mathbf{b}) \right\} \\
&= \max_{\boldsymbol{\lambda} \in \mathcal{B}_0} \left\{ \theta(\bar{\mathbf{u}}_t) - \theta(\mathbf{u}_*) + \bar{\boldsymbol{\lambda}}_t^T (A\mathbf{x}_* + B\mathbf{y}_* - b) - \boldsymbol{\lambda}^T (A\bar{\mathbf{x}}_t + B\bar{\mathbf{y}}_t - \mathbf{b}) \right\} \\
&= \max_{\boldsymbol{\lambda} \in \mathcal{B}_0} \left\{ \theta(\bar{\mathbf{u}}_t) - \theta(\mathbf{u}_*) - \boldsymbol{\lambda}^T (A\bar{\mathbf{x}}_t + B\bar{\mathbf{y}}_t - \mathbf{b}) \right\} \\
&= \theta(\bar{\mathbf{u}}_t) - \theta(\mathbf{u}_*) + \rho \|A\bar{\mathbf{x}}_t + B\bar{\mathbf{y}}_t - \mathbf{b}\|_2
\end{aligned} \tag{33}$$

Taking an expectation over (33) and using (32) we have:

$$
\begin{aligned}
&\mathbb{E}\left[\theta(\bar{\mathbf{u}}_t) - \theta(\mathbf{u}_*) + \rho\|A\bar{\mathbf{x}}_t + B\bar{\mathbf{y}}_t - \mathbf{b}\|_2\right] \\
\leq\ & \mathbb{E}\left[\frac{1}{t}\sum_{k=0}^{t-1}\left(\frac{\eta_{k+1}\|\theta_1'(\mathbf{x}_k, \boldsymbol{\xi}_{k+1})\|^2}{2} + \langle \boldsymbol{\delta}_{k+1}, \mathbf{x}_* - \mathbf{x}_k\rangle\right) + \frac{1}{t}\left(\frac{D_{\mathcal{X}}^2}{2\eta_t} + \frac{\beta}{2}D_{\mathbf{y}_*,B}^2\right)\right] \\
&+ \mathbb{E}\left[\max_{\boldsymbol{\lambda}\in\mathcal{B}_0}\left\{\frac{1}{2\beta t}\|\boldsymbol{\lambda} - \boldsymbol{\lambda}_0\|_2^2\right\}\right] \\
\leq\ & \frac{1}{t}\left(\frac{M^2}{2}\sum_{k=1}^{t}\eta_k + \frac{D_{\mathcal{X}}^2}{2\eta_t}\right) + \frac{\beta D_{\mathbf{y}_*,B}^2}{2t} + \frac{\rho^2}{2\beta t} + \frac{1}{t}\sum_{k=0}^{t-1}\mathbb{E}\left[\langle \boldsymbol{\delta}_{k+1}, \mathbf{x}_* - \mathbf{x}_k\rangle\right] \\
=\ & \frac{1}{t}\left(\frac{M^2}{2}\sum_{k=1}^{t}\eta_k + \frac{D_{\mathcal{X}}^2}{2\eta_t}\right) + \frac{\beta D_{\mathbf{y}_*,B}^2}{2t} + \frac{\rho^2}{2\beta t} \\
\leq\ & \frac{\sqrt{2}D_{\mathcal{X}}M}{\sqrt{t}} + \frac{\beta D_{\mathbf{y}_*,B}^2}{2t} + \frac{\rho^2}{2\beta t}
\end{aligned}
$$

In the second last step, we use the fact that $\mathbf{x}_k$ is independent of $\boldsymbol{\xi}_{k+1}$, hence $\mathbb{E}_{\boldsymbol{\xi}_{k+1}|\boldsymbol{\xi}_{[1:k]}}\langle \boldsymbol{\delta}_{k+1}, \mathbf{x}_* - \mathbf{x}_k\rangle = \left\langle \mathbb{E}_{\boldsymbol{\xi}_{k+1}|\boldsymbol{\xi}_{[1:k]}}\boldsymbol{\delta}_{k+1}, \mathbf{x}_* - \mathbf{x}_k\right\rangle = 0$.

(ii) From the steps in the proof of part (i), it follows that,

$$
\begin{aligned}
&\theta(\bar{\mathbf{u}}_t) - \theta(\mathbf{u}_*) + \rho\|A\bar{\mathbf{x}}_t + B\bar{\mathbf{y}}_t - \mathbf{b}\| \\
\leq\ & \frac{1}{t}\sum_{k=0}^{t-1}\frac{\eta_{k+1}\ \|\theta_1'(\mathbf{x}_k, \boldsymbol{\xi}_{k+1})\|^2}{2} + \frac{1}{t}\sum_{k=0}^{t-1}\langle \boldsymbol{\delta}_{k+1}, \mathbf{x}_* - \mathbf{x}_k\rangle + \frac{1}{t}\left(\frac{D_{\mathcal{X}}^2}{2\eta_t} + \frac{\beta}{2}D_{\mathbf{y}_*,B}^2 + \frac{\rho^2}{2\beta}\right) \\
\equiv\ & A_t + B_t + C_t
\end{aligned}
\tag{34}
$$

Note that random variables A_t and B_t are dependent on $\boldsymbol{\xi}_{[t]}$.

Claim 1. *For* $\Omega_1 > 0$,

$$
\mathit{Prob}\left(A_t \geq (1+\Omega_1)\frac{M^2}{2t}\sum_{k=1}^{t}\eta_k\right) \leq \exp\{-\Omega_1\}. \tag{35}
$$

Let $\alpha_k \equiv \frac{\eta_k}{\sum_{k=1}^{t}\eta_k}$ $\forall k = 1, \ldots, t$, then $0 \leq \alpha_k \leq 1$ and $\sum_{k=1}^{t}\alpha_k = 1$. Using the fact that $\{\boldsymbol{\delta}_k, \forall k\}$ are independent and applying Assumption 2, one has

$$
\begin{aligned}
\mathbb{E}\left[\exp\left\{\sum_{k=1}^{t}\alpha_k\|\theta_1'(\mathbf{x}_k, \boldsymbol{\xi}_{k+1})\|^2/M^2\right\}\right] &= \prod_{k=1}^{t}\mathbb{E}\left[\exp\left\{\alpha_k\|\theta_1'(\mathbf{x}_k, \boldsymbol{\xi}_{k+1})\|^2/M^2\right\}\right] \\
&\leq \prod_{k=1}^{t}\left(\mathbb{E}\left[\exp\left\{\|\theta_1'(\mathbf{x}_k, \boldsymbol{\xi}_{k+1})\|^2/M^2\right\}\right]\right)^{\alpha_k} \qquad \text{(Jensen's Inequality)} \\
&\leq \prod_{k=1}^{t}\left(\exp\{1\}\right)^{\alpha_k} = \exp\left\{\sum_{k=1}^{t}\alpha_k\right\} = \exp\{1\}
\end{aligned}
$$

Hence, by Markov's Inequality, we can get

$$
\text{Prob}\left(A_t \geq (1+\Omega_1)\frac{M^2}{2t}\sum_{k=1}^{t}\eta_k\right) \leq \exp\left\{-(1+\Omega_1)\right\}\mathbb{E}\left[\exp\left\{\sum_{k=1}^{t}\alpha_k\|\theta_1'(\mathbf{x}_k, \boldsymbol{\xi}_{k+1})\|^2/M^2\right\}\right] \leq \exp\{-\Omega_1\}.
$$

We have therefore proved Claim 1.

Claim 2. *For* $\Omega_2 > 0$,

$$
\mathit{Prob}\left(B_t > 2\Omega_2\frac{D_{\mathcal{X}}M}{\sqrt{t}}\right) \leq \exp\left\{-\frac{\Omega_2^2}{4}\right\}. \tag{36}
$$

In order to prove this claim, we adopt the following facts in Nemirovski's paper (Nemirovski et al., 2009).

Lemma 3. *Given that for all $k = 1, \ldots, t$, ζ_k is a deterministic function of $\boldsymbol{\xi}_{[k]}$ with $\mathbb{E}\left[\zeta_k | \boldsymbol{\xi}_{[k-1]}\right] = 0$ and $\mathbb{E}\left[\exp\{\zeta_k^2/\sigma_k^2\}|\boldsymbol{\xi}_{[k-1]}\right] \leq \exp\{1\}$, we have*

(a) For $\gamma \geq 0$, $\mathbb{E}\left[\exp\{\gamma\zeta_k\}|\boldsymbol{\xi}_{[k-1]}\right] \leq \exp\{\gamma^2\sigma_k^2\}, \forall k = 1, \ldots, t$

(b) Let $S_t = \sum_{k=1}^t \zeta_k$, then $Prob\{S_t > \Omega\sqrt{\sum_{k=1}^t \sigma_k^2}\} \leq \exp\left\{-\frac{\Omega^2}{4}\right\}$.

Using this result by setting $\zeta_k = \langle \boldsymbol{\delta}_k, \mathbf{x}_* - \mathbf{x}_{k-1}\rangle$, $S_t = \sum_{k=1}^t \zeta_k$, and $\sigma_k = 2D_{\mathcal{X}}M, \forall k$, we can verify that $\mathbb{E}\left[\zeta_k|\boldsymbol{\xi}_{[k-1]}\right] = 0$ and

$$\mathbb{E}\left[\exp\{\zeta_k^2/\sigma_k^2\}|\boldsymbol{\xi}_{[k-1]}\right] \leq \mathbb{E}\left[\exp\{D_{\mathcal{X}}^2\|\boldsymbol{\delta}_k\|^2/\sigma_k^2\}|\boldsymbol{\xi}_{[k-1]}\right] \leq \exp\{1\},$$

since $|\zeta_k|^2 \leq \|\mathbf{x}_* - \mathbf{x}_{k-1}\|^2\|\boldsymbol{\delta}_k\|^2 \leq D_{\mathcal{X}}^2\left(2\|\theta_1'(\mathbf{x}_k, \boldsymbol{\xi}_{k+1})\|^2 + 2M^2\right)$.

Implementing the above results, it follows that

$$\text{Prob}\left(S_t > 2\Omega_2 D_{\mathcal{X}}M\sqrt{t}\right) \leq \exp\left\{-\frac{\Omega_2^2}{4}\right\}.$$

Since $S_t = tB_t$, we have

$$\text{Prob}\left(B_t > 2\Omega_2\frac{D_{\mathcal{X}}M}{\sqrt{t}}\right) \leq \exp\left\{-\frac{\Omega_2^2}{4}\right\}$$

as desired.

Combining (34), (35) and (36), we obtain

$$\text{Prob}\left(\text{Err}_\rho(\bar{\mathbf{u}}_t) > (1+\Omega_1)\frac{M^2}{2t}\sum_{k=1}^t \eta_k + 2\Omega_2\frac{D_{\mathcal{X}}M}{\sqrt{t}} + C_t\right) \leq \exp\{-\Omega_1\} + \exp\left\{-\frac{\Omega_2}{4}\right\},$$

where $\text{Err}_\rho(\bar{\mathbf{u}}_t) \equiv \theta(\bar{\mathbf{u}}_t) - \theta(\mathbf{u}_*) + \rho\|A\bar{\mathbf{x}}_t + B\bar{\mathbf{y}}_t - \mathbf{b}\|_2$. Substituting $\Omega_1 = \Omega, \Omega_2 = 2\sqrt{\Omega}$ and plugging in $\eta_k = \frac{D_{\mathcal{X}}}{M\sqrt{2k}}$, we obtain (10) as desired. □

7.4. Proof of Theorem 2

Proof. By the strong-convexity of θ_1 we have $\forall \mathbf{x}$:

$$\begin{aligned}
\theta_1(\mathbf{x}_k) - \theta_1(\mathbf{x}) &\leq \left\langle \theta_1'(\mathbf{x}_k), \mathbf{x}_k - \mathbf{x}\right\rangle - \frac{\mu}{2}\|\mathbf{x} - \mathbf{x}_k\|^2 \\
&= \left\langle \theta_1'(\mathbf{x}_k, \boldsymbol{\xi}_{k+1}), \mathbf{x}_{k+1} - \mathbf{x}\right\rangle + \langle \boldsymbol{\delta}_{k+1}, \mathbf{x} - \mathbf{x}_k\rangle + \left\langle \theta_1'(\mathbf{x}_k, \boldsymbol{\xi}_{k+1}), \mathbf{x}_k - \mathbf{x}_{k+1}\right\rangle - \frac{\mu}{2}\|\mathbf{x} - \mathbf{x}_k\|^2.
\end{aligned}$$

Following the same derivations as in Lemma 2 and Theorem 1 (i), we have

$$\begin{aligned}
&\mathbb{E}\left[\theta(\bar{\mathbf{u}}_t) - \theta(\mathbf{u}_*) + \rho\|A\bar{\mathbf{x}}_t + B\bar{\mathbf{y}}_t - \mathbf{b}\|_2\right] \\
&\leq \mathbb{E}\left\{\frac{1}{t}\sum_{k=0}^{t-1}\left[\frac{\eta_{k+1}\|\theta_1'(\mathbf{x}_k, \boldsymbol{\xi}_{k+1})\|^2}{2} + \left(\frac{1}{2\eta_{k+1}} - \frac{\mu}{2}\right)\|\mathbf{x}_k - \mathbf{x}_*\|^2 - \frac{\|\mathbf{x}_{k+1} - \mathbf{x}_*\|^2}{2\eta_{k+1}}\right]\right\} \\
&+ \frac{\beta D_{\mathbf{y}_*,B}^2}{2t} + \mathbb{E}\left[\max_{\boldsymbol{\lambda}\in\mathcal{B}_0}\left\{\frac{1}{2\beta t}\|\boldsymbol{\lambda} - \boldsymbol{\lambda}_0\|_0^2\right\}\right] \\
&\leq \frac{M^2}{2t}\sum_{k=1}^t \frac{1}{\mu k} + \frac{1}{t}\sum_{k=0}^{t-1}\mathbb{E}\left[\frac{\mu k}{2}\|\mathbf{x}_k - \mathbf{x}_*\|^2 - \frac{\mu(k+1)}{2}\|\mathbf{x}_{k+1} - \mathbf{x}_*\|^2\right] + \frac{\beta D_{\mathbf{y}_*,B}^2}{2t} + \frac{\rho^2}{2\beta t} \\
&\leq \frac{M^2\log t}{\mu t} + \frac{\mu D_{\mathcal{X}}^2}{2t} + \frac{\beta D_{\mathbf{y}_*,B}^2}{2t} + \frac{\rho^2}{2\beta t}.
\end{aligned}$$

□

7.5. Proof of Theorem 3

Proof. The Lipschitz smoothness of θ_1 implies that $\forall k \geq 0$:

$$\begin{aligned}
\theta_1(\mathbf{x}_{k+1}) &\leq \theta_1(\mathbf{x}_k) + \langle \nabla\theta_1(\mathbf{x}_k), \mathbf{x}_{k+1} - \mathbf{x}_k\rangle + \frac{L}{2}\|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2 \\
&\overset{(3)}{=} \theta_1(\mathbf{x}_k) + \langle \nabla\theta_1(\mathbf{x}_k, \boldsymbol{\xi}_{k+1}), \mathbf{x}_{k+1} - \mathbf{x}_k\rangle - \langle \boldsymbol{\delta}_{k+1}, \mathbf{x}_{k+1} - \mathbf{x}_k\rangle + \frac{L}{2}\|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2.
\end{aligned}$$

It follows that $\forall \mathbf{x} \in \mathcal{X}$:

$$
\begin{aligned}
&\theta_1(\mathbf{x}_{k+1}) - \theta_1(\mathbf{x}) + \left\langle \mathbf{x}_{k+1} - \mathbf{x}, -A^T\boldsymbol{\lambda}_{k+1} \right\rangle \\
\leq\ & \theta_1(\mathbf{x}_k) - \theta_1(\mathbf{x}) + \langle \nabla\theta_1(\mathbf{x}_k, \boldsymbol{\xi}_{k+1}), \mathbf{x}_{k+1} - \mathbf{x}_k\rangle - \langle \boldsymbol{\delta}_{k+1}, \mathbf{x}_{k+1} - \mathbf{x}_k\rangle + \frac{L}{2}\|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2 + \left\langle \mathbf{x}_{k+1} - \mathbf{x}, -A^T\boldsymbol{\lambda}_{k+1} \right\rangle \\
=\ & \theta_1(\mathbf{x}_k) - \theta_1(\mathbf{x}) + \langle \nabla\theta_1(\mathbf{x}_k, \boldsymbol{\xi}_{k+1}), \mathbf{x} - \mathbf{x}_k\rangle - \langle \boldsymbol{\delta}_{k+1}, \mathbf{x}_{k+1} - \mathbf{x}_k\rangle + \frac{L}{2}\|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2 \\
&+ \left[\langle \nabla\theta_1(\mathbf{x}_k, \boldsymbol{\xi}_{k+1}), \mathbf{x}_{k+1} - \mathbf{x}\rangle + \left\langle \mathbf{x}_{k+1} - \mathbf{x}, -A^T\boldsymbol{\lambda}_{k+1} \right\rangle\right] \\
\leq\ & \langle \nabla\theta_1(\mathbf{x}_k), \mathbf{x}_k - \mathbf{x}\rangle + \langle \nabla\theta_1(\mathbf{x}_k, \boldsymbol{\xi}_{k+1}), \mathbf{x} - \mathbf{x}_k\rangle - \langle \boldsymbol{\delta}_{k+1}, \mathbf{x}_{k+1} - \mathbf{x}_k\rangle + \frac{L}{2}\|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2 \\
&+ \left[\langle \nabla\theta_1(\mathbf{x}_k, \boldsymbol{\xi}_{k+1}), \mathbf{x}_{k+1} - \mathbf{x}\rangle + \left\langle \mathbf{x}_{k+1} - \mathbf{x}, -A^T\boldsymbol{\lambda}_{k+1} \right\rangle\right] \\
=\ & \langle \boldsymbol{\delta}_{k+1}, \mathbf{x} - \mathbf{x}_{k+1}\rangle + \frac{L}{2}\|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2 + \left[\langle \nabla\theta_1(\mathbf{x}_k, \boldsymbol{\xi}_{k+1}), \mathbf{x}_{k+1} - \mathbf{x}\rangle + \left\langle \mathbf{x}_{k+1} - \mathbf{x}, -A^T\boldsymbol{\lambda}_{k+1} \right\rangle\right] \\
=\ & \langle \boldsymbol{\delta}_{k+1}, \mathbf{x} - \mathbf{x}_{k+1}\rangle + \frac{L}{2}\|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2 + \left\langle \mathbf{x} - \mathbf{x}_{k+1}, \beta A^T B(\mathbf{y}_k - \mathbf{y}_{k+1}) \right\rangle \\
&+ \left\langle \nabla\theta_1(\mathbf{x}_k, \boldsymbol{\xi}_{k+1}) + A^T\left[\beta(A\mathbf{x_{k+1}} + B\mathbf{y}_k - \mathbf{b}) - \boldsymbol{\lambda}_k\right], \mathbf{x}_{k+1} - \mathbf{x} \right\rangle \\
\stackrel{(24)}{\leq}\ & \frac{1}{2\eta_{k+1}}\left(\|\mathbf{x} - \mathbf{x}_k\|^2 - \|\mathbf{x} - \mathbf{x}_{k+1}\|^2\right) - \frac{1/\eta_{k+1} - L}{2}\|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2 \\
&+ \left\langle \mathbf{x} - \mathbf{x}_{k+1}, \beta A^T B(\mathbf{y}_k - \mathbf{y}_{k+1}) \right\rangle + \langle \boldsymbol{\delta}_{k+1}, \mathbf{x} - \mathbf{x}_{k+1}\rangle .
\end{aligned}
$$

The last inner product can be bounded as below using Young's inequality, given that $\eta_{k+1} \leq \frac{1}{L}$:

$$
\begin{aligned}
\langle \boldsymbol{\delta}_{k+1}, \mathbf{x} - \mathbf{x}_{k+1}\rangle &= \langle \boldsymbol{\delta}_{k+1}, \mathbf{x} - \mathbf{x}_k\rangle + \langle \boldsymbol{\delta}_{k+1}, \mathbf{x}_k - \mathbf{x}_{k+1}\rangle \\
&\leq \langle \boldsymbol{\delta}_{k+1}, \mathbf{x} - \mathbf{x}_k\rangle + \frac{1}{2\left(1/\eta_{k+1} - L\right)}\|\boldsymbol{\delta}_{k+1}\|^2 + \frac{1/\eta_{k+1} - L}{2}\|\mathbf{x}_k - \mathbf{x}_{k+1}\|^2.
\end{aligned}
$$

Combining this with inequalities (26,29) and (30), we can get a similar statement as that of Lemma 2:

$$
\begin{aligned}
&\theta(\mathbf{u}_{k+1}) - \theta(\mathbf{u}) + (\mathbf{w}_{k+1} - \mathbf{w})^T F(\mathbf{w}_{k+1}) \leq \frac{\|\boldsymbol{\delta}_{k+1}\|^2}{2(1/\eta_{k+1} - L)} \\
&+ \frac{1}{2\eta_{k+1}}\left(\|\mathbf{x}_k - \mathbf{x}\|^2 - \|\mathbf{x}_{k+1} - \mathbf{x}\|^2\right) + \frac{\beta}{2}\left(\|A\mathbf{x} + B\mathbf{y}_k - \mathbf{b}\|^2 - \|A\mathbf{x} + B\mathbf{y}_{k+1} - \mathbf{b}\|^2\right) \\
&+ \langle \boldsymbol{\delta}_{k+1}, \mathbf{x} - \mathbf{x}_k\rangle + \frac{1}{2\beta}\left(\|\boldsymbol{\lambda} - \boldsymbol{\lambda}_k\|_2^2 - \|\boldsymbol{\lambda} - \boldsymbol{\lambda}_{k+1}\|_2^2\right).
\end{aligned}
$$

The rest of the proof are essentially the same as Theorem 1 (i), except that we use the new definition of $\bar{\mathbf{u}}_k$ in (12). ☐