A Proof of Lemma 2

For simplicity, we assume that $d_1 = ... = d_p = m = d/p$. For any $s \in \mathbb{Z}^+$, we define the lower triangular matrix $D_s \in \mathbb{R}^{s\times s}$ as

$$D_s = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ 1 & 1 & 0 & \dots & 0 & 0 \\ 1 & 1 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 1 & 1 & 1 & \dots & 1 & 1 \end{bmatrix}$$

By the definition of L_j, we have

$$L_j \geq \lambda_{max}(A_j^\top A_j), \forall j.$$

Then we have

$$\begin{aligned}
&\mathcal{F}(x^{(t+1)}) - \mathcal{F}(x^*) \\
&\overset{\text{(i)}}{\leq} \langle \nabla \mathcal{L}(x^{(t+1)}), x^{(t+1)} - x^* \rangle + \mathcal{R}(x^{(t+1)}) - \mathcal{R}(x^*) \\
&\overset{\text{(ii)}}{\leq} \langle \nabla \mathcal{L}(x^{(t+1)}), x^{(t+1)} - x^* \rangle + \langle \xi^{(t+1)}, x^{(t+1)} - x^* \rangle - \frac{\mu}{2}\|x^{(t+1)} - x^*\|^2 \\
&\overset{\text{(iii)}}{\leq} \sum_{j=1}^{p} \langle \nabla_j \mathcal{L}(x^{(t+1)}) - \nabla_j \mathcal{L}\left(x^{(t)}\right), x_j^{(t+1)} - x_j^* \rangle - \sum_{j=1}^{p} L_j \langle x_j^{(t+1)} - x_j^{(t)}, x_j^{(t+1)} - x_j^* \rangle - \frac{\mu}{2}\|x^{(t+1)} - x^*\|^2 \\
&= \sum_{j=1}^{p} \left\langle \sum_{k\geq j} A_k (x_k^{(t+1)} - x_k^{(t)}), A_j(x_j^{(t+1)} - x_j^*) \right\rangle - (x^{(t+1)} - x^{(t)})^\top (\widetilde{P} \otimes I_m)(x^{(t+1)} - x^*) - \frac{\mu}{2}\|x^{(t+1)} - x^*\|^2 \\
&\leq (x^{(t+1)} - x^{(t)})^\top \widetilde{A}^\top (D_p \otimes I_m)\widetilde{A}(x^{(t+1)} - x^*) - (x^{(t+1)} - x^{(t)})^\top (\widetilde{P} \otimes I_m)(x^{(t+1)} - x^*) - \frac{\mu}{2}\|x^{(t+1)} - x^*\|^2 \\
&= (x^{(t+1)} - x^{(t)})^\top \left(\widetilde{A}^\top (D_p \otimes I_m)\widetilde{A} - \widetilde{P} \otimes I_m\right)(x^{(t+1)} - x^*) - \frac{\mu}{2}\|x^{(t+1)} - x^*\|^2 \\
&\overset{\text{(iv)}}{=} (x^{(t+1)} - x^{(t)})^\top \left(\left(A^\top A - \widetilde{A}^\top \widetilde{A}\right) \odot D_d + \widetilde{A}^\top \widetilde{A} - \widetilde{P} \otimes I_m\right)(x^{(t+1)} - x^*) - \frac{\mu}{2}\|x^{(t+1)} - x^*\|^2,
\end{aligned}$$

where (i) is from (11), (ii) is from Assumption 2, (iii) is from the optimality condition to the subproblem associated with x_j,

$$\langle \nabla_j \mathcal{L}(x^{(t)}) + L_j(x_j^{(t+1)} - x_j^{(t)}) + \xi_j^{(t+1)}, x_j - x_j^{(t+1)} \rangle \geq 0 \text{ for any } x_j \in \mathbb{R}^m,$$

and (iv) comes from the fact that

$$\widetilde{A}^\top (D_p \otimes I_m)\widetilde{A} = \left(A^\top A - \widetilde{A}^\top \widetilde{A}\right) \odot D_d + \widetilde{A}^\top \widetilde{A},$$

where $\odot$ denotes the Hadamard product and $\mathbf{1}_n \in \mathbb{R}^{n\times n}$ is a matrix with all entries as 1.

Let us define

$$B = \left(A^\top A - \widetilde{A}^\top \widetilde{A}\right) \odot D_d + \widetilde{A}^\top \widetilde{A} - \widetilde{P} \otimes I_m,$$

then we have

$$\mathcal{F}(x^{(t+1)}) - \mathcal{F}(x^*) \leq (x^{(t+1)} - x^{(t)})^\top B (x^{(t+1)} - x^*) - \frac{\mu}{2}\|x^{(t+1)} - x^*\|^2, \tag{28}$$

Maximizing R.H.S. of the above inequality over x^*, we obtain

$$-\mu(x^* - x^{(t+1)}) - B^\top (x^{(t+1)} - x^{(t)}) = 0.$$

which implies

$$x^* = -\frac{B^\top(x^{(t+1)} - x^{(t)})}{\mu} + x^{(t+1)}. \tag{29}$$

Plugging (29) into (28), we obtain

$$\begin{aligned}
\mathcal{F}(x^{(t+1)}) - \mathcal{F}(x^*) &\le \frac{1}{2\mu}\|B(x^{(t+1)} - x^{(t)})\|^2 \le \frac{\lambda_{\max}^2(B)}{2\mu}\|x^{(t+1)} - x^{(t)}\|^2 \\
&\overset{\text{(i)}}{\le} \frac{\lambda_{\max}^2\left(\left(A^\top A - \widetilde{A}^\top \widetilde{A}\right) \odot D_d\right)}{2\mu}\|x^{(t+1)} - x^{(t)}\|^2 \\
&\overset{\text{(ii)}}{\le} \frac{\lambda_{\max}^2\left(A^\top A - \widetilde{A}^\top \widetilde{A}\right)\left(1 + \frac{1}{\pi} + \frac{\log(d)}{\pi}\right)^2}{2\mu}\|x^{(t+1)} - x^{(t)}\|^2 \\
&\overset{\text{(iii)}}{\le} \frac{\lambda_{\max}^2\left(A^\top A - \widetilde{A}^\top \widetilde{A}\right)\log^2(2d)}{2\mu}\|x^{(t+1)} - x^{(t)}\|^2 \\
&\overset{\text{(iv)}}{\le} \frac{L^2 \log^2(2d)}{2\mu}\|x^{(t+1)} - x^{(t)}\|^2,
\end{aligned}$$

where (i) comes from (12), which indicates that $\lambda_{\max}(\widetilde{A}^\top \widetilde{A} - \widetilde{P} \otimes I_m) \le 0$, (iii) is true if $d \ge 3$, and (iv) comes from $d \le p \cdot d_{\max}$ and the fact that

$$\lambda_{\max}\left(A^\top A - \widetilde{A}^\top \widetilde{A}\right) \le \lambda_{\max}\left(A^\top A\right) + \lambda_{\max}\left(-\widetilde{A}^\top \widetilde{A}\right) \le \lambda_{\max}\left(A^\top A\right) \le L.$$

Inequality (ii) follows from the result on the spectral norm of the triangular truncation operator in (Angelos et al., 1992). More specifically, let us define

$$L_d = \max\left\{\frac{\|A \odot D_d\|}{\|A\|} : A \in \mathbb{R}^{d\times d}, A \ne \mathbf{0}\right\}.$$

Then we have

$$\left|\frac{L_d}{\log d} - \frac{1}{\pi}\right| \le \frac{\left(1 + \frac{1}{\pi}\right)}{\log d}.$$

The final claim holds by the fact that $d \le p \cdot d_{\max}$.