This appendix is organized as following.

- Section A provides missing proofs of the theoretical results in Section 3.
- Section B provides missing proofs of the theoretical results in Section 4.
- Sectin C provides more details of experiments.

Appendix A. Missing Proofs of Section 3

A.1. Proof of Lemma 3

Proof Substituting (12a) into (9a), we obtain

$$\left\{\widehat{f}_\rho(\mathbf{w}_{k+1}) + \frac{1}{2}\|\mathbf{w}_{k+1} + \mathbf{s}_k - \mathbf{c} - \mathbf{u}_k\|^2\right\} - \min_{\mathbf{w}}\left\{\widehat{f}_\rho(\mathbf{w}) + \frac{1}{2}\|\mathbf{w} + \mathbf{s}_k - \mathbf{c} - \mathbf{u}_k\|^2\right\} \leq \epsilon_{k+1}/\rho.$$

Applying strong convexity of the objective, the left-hand side can be relaxed as

$$\begin{aligned}\left\{\widehat{f}_\rho(\mathbf{w}_{k+1}) + \frac{1}{2}\|\mathbf{w}_{k+1} + \mathbf{s}_k - \mathbf{c} - \mathbf{u}_k\|^2\right\} - \min_{\mathbf{w}}\Big\{\widehat{f}_\rho(\mathbf{w}_{k+1}) + \frac{1}{2}\|\mathbf{w}_{k+1} + \mathbf{s}_k - \mathbf{c} - \mathbf{u}_k\|^2 \\ + \left\langle \nabla\widehat{f}_\rho(\mathbf{w}_{k+1}) + \mathbf{w}_{k+1} + \mathbf{s}_k - \mathbf{c} - \mathbf{u}_k, \mathbf{w} - \mathbf{w}_{k+1}\right\rangle + \frac{\widehat{m}+\rho}{2\rho}\|\mathbf{w} - \mathbf{w}_{k+1}\|^2\Big\} \leq \epsilon_{k+1}/\rho.\end{aligned}$$

It is equivalent to

$$-\min_{\mathbf{w}}\left\{\left\langle \nabla\widehat{f}_\rho(\mathbf{w}_{k+1}) + \mathbf{w}_{k+1} + \mathbf{s}_k - \mathbf{c} - \mathbf{u}_k, \mathbf{w} - \mathbf{w}_{k+1}\right\rangle + \frac{\widehat{m}+\rho}{2\rho}\|\mathbf{w} - \mathbf{w}_{k+1}\|^2\right\} \leq \epsilon_{k+1}/\rho. \quad \text{(A30)}$$

Taking the optimum for $\mathbf{w}$, we obtain

$$\left\|\nabla\widehat{f}_\rho(\mathbf{w}_{k+1}) + \mathbf{w}_{k+1} + \mathbf{s}_k - \mathbf{c} - \mathbf{u}_k\right\|^2 \leq \frac{2(\rho+\widehat{m})\epsilon_{k+1}}{\rho^2}.$$

This implies that there exists $\boldsymbol{\eta}_{k+1} \in \mathbb{R}^p$ with $\|\boldsymbol{\eta}_{k+1}\|^2 \leq 2(\rho+\widehat{m})\epsilon_{k+1}/\rho^2$ such that

$$\nabla\widehat{f}_\rho(\mathbf{w}_{k+1}) + \mathbf{w}_{k+1} + \mathbf{s}_k - \mathbf{c} - \mathbf{u}_k + \boldsymbol{\eta}_{k+1} = \mathbf{0}.$$

Substituting $\boldsymbol{\beta}_{k+1} = \nabla\widehat{f}_\rho(\mathbf{w}_{k+1})$, it becomes

$$\mathbf{w}_{k+1} = -\mathbf{s}_k + \mathbf{u}_k - \left(\boldsymbol{\beta}_{k+1} + \boldsymbol{\eta}_{k+1}\right) + \mathbf{c}. \quad \text{(A31)}$$

Substituting (12b) into (9b),

$$\begin{aligned}&\left\{\widehat{g}_\rho(\mathbf{s}_{k+1}) + \frac{1}{2}\|\alpha\mathbf{w}_{k+1} - (1-\alpha)\mathbf{s}_k + \mathbf{s}_{k+1} - \alpha\mathbf{c} - \mathbf{u}_k\|^2\right\} \\ &- \min_{\mathbf{s}}\left\{\widehat{g}_\rho(\mathbf{s}) + \frac{1}{2}\|\alpha\mathbf{w}_{k+1} - (1-\alpha)\mathbf{s}_k + \mathbf{s} - \alpha\mathbf{c} - \mathbf{u}_k\|^2\right\} \leq \delta_{k+1}/\rho.\end{aligned}$$

Applying strong convexity of the objective, the left-hand side can be relaxed as

$$\left\{\widehat{g}_\rho(\mathbf{s}_{k+1}) + \frac{1}{2}\|\alpha\mathbf{w}_{k+1} - (1-\alpha)\mathbf{s}_k + \mathbf{s}_{k+1} - \alpha\mathbf{c} - \mathbf{u}_k\|^2\right\}$$

$$-\min_{\mathbf{s}}\Big\{\frac{1}{2}\|\alpha\mathbf{w}_{k+1}-(1-\alpha)\mathbf{s}_k+\mathbf{s}_{k+1}-\alpha\mathbf{c}-\mathbf{u}_k\|^2+\widehat{g}_\rho(\mathbf{s}_{k+1})$$
$$+\big\langle\boldsymbol{\gamma}_{k+1}+\alpha\mathbf{w}_{k+1}-(1-\alpha)\mathbf{s}_k+\mathbf{s}_{k+1}-\alpha\mathbf{c}-\mathbf{u}_k,\mathbf{s}-\mathbf{s}_{k+1}\big\rangle+\frac{1}{2}\|\mathbf{s}-\mathbf{s}_{k+1}\|^2\Big\}\le\delta_{k+1}/\rho.$$

It is equivalent to

$$-\min_{\mathbf{s}}\Big\{\big\langle\boldsymbol{\gamma}_{k+1}+\alpha\mathbf{w}_{k+1}-(1-\alpha)\mathbf{s}_k+\mathbf{s}_{k+1}-\alpha\mathbf{c}-\mathbf{u}_k,\mathbf{s}-\mathbf{s}_{k+1}\big\rangle+\frac{1}{2}\|\mathbf{s}-\mathbf{s}_{k+1}\|^2\Big\}\le\delta_{k+1}/\rho.$$

Taking the optimum for $\mathbf{s}$, we obtain

$$\big\|\boldsymbol{\gamma}_{k+1}+\alpha\mathbf{w}_{k+1}-(1-\alpha)\mathbf{s}_k+\mathbf{s}_{k+1}-\alpha\mathbf{c}-\mathbf{u}_k\big\|^2\le\frac{2\delta_{k+1}}{\rho}.$$

This implies that there exists $\boldsymbol{\zeta}_{k+1}\in\mathbb{R}^p$ with $\|\boldsymbol{\zeta}_{k+1}\|^2\le 2\delta_{k+1}/\rho$ such that

$$\boldsymbol{\gamma}_{k+1}+\alpha\mathbf{w}_{k+1}-(1-\alpha)\mathbf{s}_k+\mathbf{s}_{k+1}-\alpha\mathbf{c}-\mathbf{u}_k+\boldsymbol{\zeta}_{k+1}=\mathbf{0}. \tag{A32}$$

It follows then that

$$\mathbf{s}_{k+1}=-\alpha\mathbf{w}_{k+1}+(1-\alpha)\mathbf{s}_k+\mathbf{u}_k-(\boldsymbol{\gamma}_{k+1}+\boldsymbol{\zeta}_{k+1})+\alpha\mathbf{c}.$$

Substituting (A31), we obtain

$$\mathbf{s}_{k+1}=\mathbf{s}_k+(1-\alpha)\mathbf{u}_k+\alpha\big(\boldsymbol{\beta}_{k+1}+\boldsymbol{\eta}_{k+1}\big)-(\boldsymbol{\gamma}_{k+1}+\boldsymbol{\zeta}_{k+1}). \tag{A33}$$

Combining (12c) and (A32), we obtain

$$\mathbf{u}_{k+1}=\boldsymbol{\gamma}_{k+1}+\boldsymbol{\zeta}_{k+1}. \tag{A34}$$

Given (A31), (A33) and (A34), it is straightforward to show

$$\begin{aligned}\boldsymbol{\xi}_{k+1}&=\big(\widehat{\mathbf{A}}\otimes\mathbf{I}_p\big)\boldsymbol{\xi}_k+\big(\widehat{\mathbf{B}}\otimes\mathbf{I}_p\big)\mathbf{v}_k,\\ \mathbf{y}_k^1&=\big(\widehat{\mathbf{C}}^1\otimes\mathbf{I}_p\big)\boldsymbol{\xi}_k+\big(\widehat{\mathbf{D}}^1\otimes\mathbf{I}_p\big)\mathbf{v}_k,\\ \mathbf{y}_k^2&=\big(\widehat{\mathbf{C}}^2\otimes\mathbf{I}_p\big)\boldsymbol{\xi}_k+\big(\widehat{\mathbf{D}}^2\otimes\mathbf{I}_p\big)\mathbf{v}_k.\end{aligned}$$

This completes the proof. ∎

A.2. Proof of Lemma 4

Before starting our main proof, we first introduce the following lemma.

Lemma A12 *Under the same setting as Lemma 3, the following inequality holds for* $\forall k\ge 0$*:*

$$\|a(\mathbf{w}_{k+1}-\mathbf{w}_\star)+b(\boldsymbol{\beta}_{k+1}-\boldsymbol{\beta}_\star)\|\le\max(|a|,|b|)\big(\|\boldsymbol{\xi}_k-\boldsymbol{\xi}_\star\|+\|\boldsymbol{\eta}_{k+1}\|\big). \tag{A35}$$

Proof It can be proved by applying Lemma 1 and the definition of our dynamical system.

$$\begin{aligned}
&\max(a^2,b^2)\|\mathbf{w}_{k+1}-\mathbf{w}_\star+\boldsymbol{\beta}_{k+1}-\boldsymbol{\beta}_\star\|^2-\|a(\mathbf{w}_{k+1}-\mathbf{w}_\star)+b(\boldsymbol{\beta}_{k+1}-\boldsymbol{\beta}_\star)\|^2\\
=&\big(\max(a^2,b^2)-a^2\big)\|\mathbf{w}_{k+1}-\mathbf{w}_\star\|^2+2\big(\max(a^2,b^2)-ab\big)\langle\mathbf{w}_{k+1}-\mathbf{w}_\star,\boldsymbol{\beta}_{k+1}-\boldsymbol{\beta}_\star\rangle\\
&+\big(\max(a^2,b^2)-b^2\big)\|\boldsymbol{\beta}_{k+1}-\boldsymbol{\beta}_\star\|^2\\
\geq&\, 2\big(\max(a^2,b^2)-ab\big)\langle\mathbf{w}_{k+1}-\mathbf{w}_\star,\boldsymbol{\beta}_{k+1}-\boldsymbol{\beta}_\star\rangle\geq 0,
\end{aligned}$$

where the last inequality follows from Lemma 1. Thus,

$$\|a(\mathbf{w}_{k+1}-\mathbf{w}_\star)+b(\boldsymbol{\beta}_{k+1}-\boldsymbol{\beta}_\star)\|\leq\max(|a|,|b|)\|(\mathbf{w}_{k+1}-\mathbf{w}_\star)+(\boldsymbol{\beta}_{k+1}-\boldsymbol{\beta}_\star)\|. \tag{A36}$$

Applying the definition of our dynamical system,

$$\begin{aligned}
&(\mathbf{w}_{k+1}-\mathbf{w}_\star)+(\boldsymbol{\beta}_{k+1}-\boldsymbol{\beta}_\star)\\
=&\,-\mathbf{s}_k+\mathbf{u}_k-\big(\boldsymbol{\beta}_{k+1}+\boldsymbol{\eta}_{k+1}\big)+\mathbf{c}-\big(-\mathbf{s}_\star+\mathbf{u}_\star-\boldsymbol{\beta}_\star+\mathbf{c}\big)+(\boldsymbol{\beta}_{k+1}-\boldsymbol{\beta}_\star)\\
=&\,-(\mathbf{s}_k-\mathbf{s}_\star)+(\mathbf{u}_k-\mathbf{u}_\star)-\boldsymbol{\eta}_{k+1}.
\end{aligned}$$

Thus, $\|(\mathbf{w}_{k+1}-\mathbf{w}_\star)+(\boldsymbol{\beta}_{k+1}-\boldsymbol{\beta}_\star)\|$ becomes

$$\begin{aligned}
\|(\mathbf{w}_{k+1}-\mathbf{w}_\star)+(\boldsymbol{\beta}_{k+1}-\boldsymbol{\beta}_\star)\|&\leq\|(\mathbf{s}_k-\mathbf{s}_\star)-(\mathbf{u}_k-\mathbf{u}_\star)\|+\|\boldsymbol{\eta}_{k+1}\|\\
&=\|(\mathbf{s}_k-\mathbf{s}_\star)-(\boldsymbol{\gamma}_k-\boldsymbol{\gamma}_\star)\|+\|\boldsymbol{\eta}_{k+1}\|\\
&\leq\sqrt{\|\mathbf{s}_k-\mathbf{s}_\star\|^2+\|\boldsymbol{\gamma}_k-\boldsymbol{\gamma}_\star\|^2}+\|\boldsymbol{\eta}_{k+1}\|\\
&=\left\|\begin{bmatrix}\mathbf{s}_k-\mathbf{s}_\star\\ \boldsymbol{\gamma}_k-\boldsymbol{\gamma}_\star\end{bmatrix}\right\|+\|\boldsymbol{\eta}_{k+1}\|\\
&=\left\|\begin{bmatrix}\mathbf{s}_k-\mathbf{s}_\star\\ \mathbf{u}_k-\mathbf{u}_\star\end{bmatrix}\right\|+\|\boldsymbol{\eta}_{k+1}\|\\
&=\|\boldsymbol{\xi}_k-\boldsymbol{\xi}_\star\|+\|\boldsymbol{\eta}_{k+1}\|,
\end{aligned}$$

where the second inequality follows from Lemma 1. Substituting it into (A36), we obtain

$$\|a(\mathbf{w}_{k+1}-\mathbf{w}_\star)+b(\boldsymbol{\beta}_{k+1}-\boldsymbol{\beta}_\star)\|\leq\max(|a|,|b|)\big(\|\boldsymbol{\xi}_k-\boldsymbol{\xi}_\star\|+\|\boldsymbol{\eta}_{k+1}\|\big).$$

This completes the proof. ■

Next, we start to prove Lemma 4.

A.3. proof of Lemma 4

Proof By the fact that $\boldsymbol{\xi}_\star$ is a fixed point of (14), $\big(\boldsymbol{\xi}_{k+1}-\boldsymbol{\xi}_\star\big)$ can be rewritten as

$$\boldsymbol{\xi}_{k+1}-\boldsymbol{\xi}_\star=\big((\widehat{\mathbf{A}}\otimes\mathbf{I}_p)\boldsymbol{\xi}_k+(\widehat{\mathbf{B}}\otimes\mathbf{I}_p)\mathbf{v}_k\big)-\big((\widehat{\mathbf{A}}\otimes\mathbf{I}_p)\boldsymbol{\xi}_\star+(\widehat{\mathbf{B}}\otimes\mathbf{I}_p)\mathbf{v}_\star\big)=\begin{bmatrix}\widehat{\mathbf{A}}\otimes\mathbf{I}_p & \widehat{\mathbf{B}}\otimes\mathbf{I}_p\end{bmatrix}\begin{bmatrix}\boldsymbol{\xi}_k-\boldsymbol{\xi}_\star\\ \mathbf{v}_k-\mathbf{v}_\star\end{bmatrix}.$$

Thus,

$$V_{\mathbf{P}}(\boldsymbol{\xi}_{k+1})-\tau^2V_{\mathbf{P}}(\boldsymbol{\xi}_k)=\begin{bmatrix}\boldsymbol{\xi}_k-\boldsymbol{\xi}_\star\\ \mathbf{v}_k-\mathbf{v}_\star\end{bmatrix}^\top\left(\begin{bmatrix}\widehat{\mathbf{A}}^\top\widehat{\mathbf{P}}\widehat{\mathbf{A}}-\tau^2\widehat{\mathbf{P}} & \widehat{\mathbf{A}}^\top\widehat{\mathbf{P}}\widehat{\mathbf{B}}\\ \widehat{\mathbf{B}}^\top\widehat{\mathbf{P}}\widehat{\mathbf{A}} & \widehat{\mathbf{B}}^\top\widehat{\mathbf{P}}\widehat{\mathbf{B}}\end{bmatrix}\otimes\mathbf{I}_p\right)\begin{bmatrix}\boldsymbol{\xi}_k-\boldsymbol{\xi}_\star\\ \mathbf{v}_k-\mathbf{v}_\star\end{bmatrix}.$$

Applying (19), it becomes

$$V_{\mathbf{P}}(\boldsymbol{\xi}_{k+1}) - \tau^2 V_{\mathbf{P}}(\boldsymbol{\xi}_k) \le \begin{bmatrix} \boldsymbol{\xi}_k - \boldsymbol{\xi}_\star \\ \mathbf{v}_k - \mathbf{v}_\star \end{bmatrix}^\top \left(\left(\begin{bmatrix} \widehat{\mathbf{C}}^1 & \widehat{\mathbf{D}}^1 \\ \widehat{\mathbf{C}}^2 & \widehat{\mathbf{D}}^2 \end{bmatrix}^\top \begin{bmatrix} \lambda^1 \widehat{\mathbf{M}}^1 & \mathbf{0} \\ \mathbf{0} & \lambda^2 \widehat{\mathbf{M}}^2 \end{bmatrix} \begin{bmatrix} \widehat{\mathbf{C}}^1 & \widehat{\mathbf{D}}^1 \\ \widehat{\mathbf{C}}^2 & \widehat{\mathbf{D}}^2 \end{bmatrix} \right) \otimes \mathbf{I}_p \right) \begin{bmatrix} \boldsymbol{\xi}_k - \boldsymbol{\xi}_\star \\ \mathbf{v}_k - \mathbf{v}_\star \end{bmatrix}$$
$$= \begin{bmatrix} \mathbf{y}_k^1 - \mathbf{y}_\star^1 \\ \mathbf{y}_k^2 - \mathbf{y}_\star^2 \end{bmatrix}^\top \left(\begin{bmatrix} \lambda^1 \widehat{\mathbf{M}}^1 & \mathbf{0}_2 \\ \mathbf{0}_2 & \lambda^2 \widehat{\mathbf{M}}^2 \end{bmatrix} \otimes \mathbf{I}_p \right) \begin{bmatrix} \mathbf{y}_k^1 - \mathbf{y}_\star^1 \\ \mathbf{y}_k^2 - \mathbf{y}_\star^2 \end{bmatrix}$$
$$= \lambda^1 \left(\mathbf{y}_k^1 - \mathbf{y}_\star^1\right)^\top \left(\widehat{\mathbf{M}}^1 \otimes \mathbf{I}_p\right) \left(\mathbf{y}_k^1 - \mathbf{y}_\star^1\right) + \lambda^2 \left(\mathbf{y}_k^2 - \mathbf{y}_\star^2\right)^\top \left(\widehat{\mathbf{M}}^2 \otimes \mathbf{I}_p\right) \left(\mathbf{y}_k^2 - \mathbf{y}_\star^2\right).$$

Substituting $\widehat{\mathbf{M}}^1$ and $\widehat{\mathbf{M}}^2$, we obtain

$$V_{\mathbf{P}}(\boldsymbol{\xi}_{k+1}) - \tau^2 V_{\mathbf{P}}(\boldsymbol{\xi}_k) \le \lambda^1 \begin{bmatrix} \mathbf{w}_{k+1} - \mathbf{w}_\star \\ \boldsymbol{\beta}_{k+1} - \boldsymbol{\beta}_\star \end{bmatrix}^\top \left(\widehat{\mathbf{M}}^1 \otimes \mathbf{I}_p\right) \begin{bmatrix} \mathbf{w}_{k+1} - \mathbf{w}_\star \\ \boldsymbol{\beta}_{k+1} - \boldsymbol{\beta}_\star \end{bmatrix} - 2\lambda^2 \langle \mathbf{s}_{k+1} - \mathbf{s}_\star, \boldsymbol{\gamma}_{k+1} - \boldsymbol{\gamma}_\star \rangle$$
$$+ 2\lambda^1 \left\langle \boldsymbol{\eta}_{k+1}, \widehat{M}_{12}^1 (\mathbf{w}_{k+1} - \mathbf{w}_\star) + \widehat{M}_{22}^1 (\boldsymbol{\beta}_{k+1} - \boldsymbol{\beta}_\star) \right\rangle + \lambda^1 \widehat{M}_{22}^1 \|\boldsymbol{\eta}_{k+1}\|^2 - 2\lambda^2 \langle \mathbf{s}_{k+1} - \mathbf{s}_\star, \boldsymbol{\zeta}_{k+1} \rangle.$$

Applying Lemmas 1 and 2, it becomes

$$V_{\mathbf{P}}(\boldsymbol{\xi}_{k+1}) - \tau^2 V_{\mathbf{P}}(\boldsymbol{\xi}_k) \le 2\lambda^1 \left\langle \boldsymbol{\eta}_{k+1}, \widehat{M}_{12}^1 (\mathbf{w}_{k+1} - \mathbf{w}_\star) + \widehat{M}_{22}^1 (\boldsymbol{\beta}_{k+1} - \boldsymbol{\beta}_\star) \right\rangle + \lambda^1 \widehat{M}_{22}^1 \|\boldsymbol{\eta}_{k+1}\|^2$$
$$- 2\lambda^2 \langle \mathbf{s}_{k+1} - \mathbf{s}_\star, \boldsymbol{\zeta}_{k+1} \rangle.$$

The right-hand side can be further relaxed as

$$V_{\mathbf{P}}(\boldsymbol{\xi}_{k+1}) - \tau^2 V_{\mathbf{P}}(\boldsymbol{\xi}_k)$$
$$\le 2\lambda^1 \left\langle \boldsymbol{\eta}_{k+1}, \widehat{M}_{12}^1 (\mathbf{w}_{k+1} - \mathbf{w}_\star) + \widehat{M}_{22}^1 (\boldsymbol{\beta}_{k+1} - \boldsymbol{\beta}_\star) \right\rangle + \lambda^1 \widehat{M}_{22}^1 \|\boldsymbol{\eta}_{k+1}\|^2 - 2\lambda^2 \langle \mathbf{s}_{k+1} - \mathbf{s}_\star, \boldsymbol{\zeta}_{k+1} \rangle$$
$$\le 2\lambda^1 \left\| \widehat{M}_{12}^1 (\mathbf{w}_{k+1} - \mathbf{w}_\star) + 2(\boldsymbol{\beta}_{k+1} - \boldsymbol{\beta}_\star) \right\| \|\boldsymbol{\eta}_{k+1}\| + 2\lambda^1 \|\boldsymbol{\eta}_{k+1}\|^2 + 2\lambda^2 \|\mathbf{s}_{k+1} - \mathbf{s}_\star\| \|\boldsymbol{\zeta}_{k+1}\|.$$

Applying Lemma A12,

$$V_{\mathbf{P}}(\boldsymbol{\xi}_{k+1}) - \tau^2 V_{\mathbf{P}}(\boldsymbol{\xi}_k)$$
$$\le 2\lambda^1 \max\left(2, |\widehat{M}_{12}^1|\right) \left(\|\boldsymbol{\xi}_k - \boldsymbol{\xi}_\star\| + \|\boldsymbol{\eta}_{k+1}\|\right) \|\boldsymbol{\eta}_{k+1}\| + 2\lambda^1 \|\boldsymbol{\eta}_{k+1}\|^2 + 2\lambda^2 \|\boldsymbol{\xi}_{k+1} - \boldsymbol{\xi}_\star\| \|\boldsymbol{\zeta}_{k+1}\|$$
$$\le 2\lambda^1 \max\left(2, |\widehat{M}_{12}^1|\right) \|\boldsymbol{\eta}_{k+1}\| \|\boldsymbol{\xi}_k - \boldsymbol{\xi}_\star\| + 2\lambda^1 \left(1 + \max\left(2, |\widehat{M}_{12}^1|\right)\right) \|\boldsymbol{\eta}_{k+1}\|^2 + 2\lambda^2 \|\boldsymbol{\xi}_{k+1} - \boldsymbol{\xi}_\star\| \|\boldsymbol{\zeta}_{k+1}\|.$$

It can be rewritten as

$$V_{\mathbf{P}}(\boldsymbol{\xi}_{k+1}) - \tau^2 V_{\mathbf{P}}(\boldsymbol{\xi}_k) \le \frac{2\lambda^1 \max(2\rho, \widehat{m} + \widehat{L})}{\rho} \|\boldsymbol{\xi}_k - \boldsymbol{\xi}_\star\| \|\boldsymbol{\eta}_{k+1}\| + \frac{2\lambda^1 \max(3\rho, \rho + \widehat{m} + \widehat{L})}{\rho} \|\boldsymbol{\eta}_{k+1}\|^2$$
$$+ 2\lambda^2 \|\boldsymbol{\xi}_{k+1} - \boldsymbol{\xi}_\star\| \|\boldsymbol{\zeta}_{k+1}\|.$$

Substituting $\|\boldsymbol{\eta}_{k+1}\|^2 \le 2(\rho + \widehat{m})\epsilon_{k+1}/\rho^2$ and $\|\boldsymbol{\xi}_{k+1} - \boldsymbol{\xi}_\star\|^2 \le 2\delta_{k+1}/\rho$,

$$V_{\mathbf{P}}(\boldsymbol{\xi}_{k+1}) - \tau^2 V_{\mathbf{P}}(\boldsymbol{\xi}_k) \le \frac{2\sqrt{2}\lambda^1 \sqrt{\rho + \widehat{m}} \max(2\rho, \widehat{m} + \widehat{L})}{\rho^2} \|\boldsymbol{\xi}_k - \boldsymbol{\xi}_\star\| \sqrt{\epsilon_{k+1}} + 2\lambda^2 \sqrt{\frac{2}{\rho}} \|\boldsymbol{\xi}_{k+1} - \boldsymbol{\xi}_\star\| \sqrt{\delta_{k+1}}$$

$$+\frac{4\lambda^1\big(\rho+\widehat{m}\big)\max(3\rho,\rho+\widehat{m}+\widehat{L})}{\rho^3}\epsilon_{k+1}.$$

Substituting $\widehat{\theta}, \widetilde{\theta}$ and $\bar{\theta}$, it becomes

$$V_{\mathbf{P}}(\boldsymbol{\xi}_{k+1})-\tau^2 V_{\mathbf{P}}(\boldsymbol{\xi}_k)\leq \widetilde{\theta}\epsilon_{k+1}+\widehat{\theta}\|\boldsymbol{\xi}_k-\boldsymbol{\xi}_\star\|\sqrt{\epsilon_{k+1}}+\bar{\theta}\|\boldsymbol{\xi}_{k+1}-\boldsymbol{\xi}_\star\|\sqrt{\delta_{k+1}}.$$

This completes the proof. ■

A.4. Proof of Theorem 7

Following is a useful lemma on non-negative sequences that will be used in the analysis of inexact over-relaxed ADMM.

Lemma A13 *Assume that an increasing sequence $\{S_k\}_{k\geq 0}$, three non-negative sequences $\{\widehat{\lambda}_k\}_{k\geq 0}, \{\overline{\lambda}_k\}_{k\geq 0}$ and $\{\beta_k\}_{k\geq 0}$ satisfy $S_0\geq \beta_0^2$ and*

$$\beta_T^2\leq S_T+\sum_{k=1}^{T}\widehat{\lambda}_k\beta_{k-1}+\sum_{k=1}^{T}\overline{\lambda}_k\beta_k, \forall T\geq 1.$$

Then, $\forall T\geq 0$:

$$\beta_T\leq \frac{1}{2}\sum_{k=1}^{T}\Big(\widehat{\lambda}_k+\overline{\lambda}_k\Big)+\left(S_T+\left(\frac{1}{2}\sum_{k=1}^{T}\Big(\widehat{\lambda}_k+\overline{\lambda}_k\Big)\right)^2\right)^{1/2}. \tag{A37}$$

$$\beta_T^2\leq S_T+\sum_{k=1}^{T}\widehat{\lambda}_k\beta_{k-1}+\sum_{k=1}^{T}\overline{\lambda}_k\beta_k\leq\left(\sqrt{S_T}+\sum_{k=1}^{T}\Big(\widehat{\lambda}_k+\overline{\lambda}_k\Big)\right)^2. \tag{A38}$$

The proof of Lemma A13 is provided in Section A.4.1.

With Lemma A13 at hand, we are ready to prove Theorem 7.

Proof For convenience, we define $E_{k+1}\stackrel{\text{def}}{=}\widetilde{\theta}\epsilon_{k+1}+\widehat{\theta}\|\boldsymbol{\xi}_k-\boldsymbol{\xi}_\star\|\sqrt{\epsilon_{k+1}}+\bar{\theta}\|\boldsymbol{\xi}_{k+1}-\boldsymbol{\xi}_\star\|\sqrt{\delta_{k+1}}, \forall k\geq 0$. Then, the result of Lemma 4 can be rewritten as

$$V_{\mathbf{P}}(\boldsymbol{\xi}_{k+1})\leq\tau^2 V_{\mathbf{P}}(\boldsymbol{\xi}_k)+E_{k+1}.$$

Applying the relationship from $k=0$ to $k=T-1$ recursively, we obtain

$$V_{\mathbf{P}}(\boldsymbol{\xi}_T)\leq\tau^{2T}V_{\mathbf{P}}(\boldsymbol{\xi}_0)+\sum_{k=1}^{T}\tau^{2(T-k)}E_k.$$

Substituting E_k, it becomes

$$V_{\mathbf{P}}(\boldsymbol{\xi}_T)\leq\tau^{2T}V_{\mathbf{P}}(\boldsymbol{\xi}_0)+\sum_{k=1}^{T}\tau^{2(T-k)}\Big(\widetilde{\theta}\epsilon_k+\widehat{\theta}\|\boldsymbol{\xi}_{k-1}-\boldsymbol{\xi}_\star\|\sqrt{\epsilon_k}+\bar{\theta}\|\boldsymbol{\xi}_k-\boldsymbol{\xi}_\star\|\sqrt{\delta_k}\Big).$$

Multiplying both sides with τ^{-2T},

$$\tau^{-2T} V_{\mathbf{P}}(\boldsymbol{\xi}_T) \leq V_{\mathbf{P}}(\boldsymbol{\xi}_0) + \widetilde{\theta} \sum_{k=1}^{T} \tau^{-2k} \epsilon_k + \widehat{\theta} \sum_{k=1}^{T} \tau^{-2k} \sqrt{\epsilon_k} \|\boldsymbol{\xi}_{k-1} - \boldsymbol{\xi}_\star\| + \bar{\theta} \sum_{k=1}^{T} \tau^{-2k} \sqrt{\delta_k} \|\boldsymbol{\xi}_k - \boldsymbol{\xi}_\star\|. \tag{A39}$$

By further relaxing the right-hand side,

$$\tau^{-2T} \sigma_{\mathbf{P}}^{\min} \|\boldsymbol{\xi}_T - \boldsymbol{\xi}_\star\| \leq V_{\mathbf{P}}(\boldsymbol{\xi}_0) + \widetilde{\theta} \sum_{k=1}^{T} \tau^{-2k} \epsilon_k + \widehat{\theta} \sum_{k=1}^{T} \tau^{-2k} \sqrt{\epsilon_k} \|\boldsymbol{\xi}_{k-1} - \boldsymbol{\xi}_\star\| + \bar{\theta} \sum_{k=1}^{T} \tau^{-2k} \sqrt{\delta_k} \|\boldsymbol{\xi}_k - \boldsymbol{\xi}_\star\|.$$

It is equivalent to

$$\tau^{-2T} \|\boldsymbol{\xi}_T - \boldsymbol{\xi}_\star\|^2 \leq \frac{1}{\sigma_{\mathbf{P}}^{\min}} \left\{ V_{\mathbf{P}}(\boldsymbol{\xi}_0) + \widetilde{\theta} \sum_{k=1}^{T} \tau^{-2k} \epsilon_k \right\} + \frac{\widehat{\theta}}{\sigma_{\mathbf{P}}^{\min}} \sum_{k=1}^{T} \tau^{-2k} \sqrt{\epsilon_k} \|\boldsymbol{\xi}_{k-1} - \boldsymbol{\xi}_\star\| + \frac{\bar{\theta}}{\sigma_{\mathbf{P}}^{\min}} \sum_{k=1}^{T} \tau^{-2k} \sqrt{\delta_k} \|\boldsymbol{\xi}_k - \boldsymbol{\xi}_\star\|.$$

It can be rewritten as

$$\begin{aligned} \left(\tau^{-T} \|\boldsymbol{\xi}_T - \boldsymbol{\xi}_\star\|\right)^2 \leq{}& \frac{1}{\sigma_{\mathbf{P}}^{\min}} \left\{ V_{\mathbf{P}}(\boldsymbol{\xi}_0) + \widetilde{\theta} \sum_{k=1}^{T} \tau^{-2k} \epsilon_k \right\} + \sum_{k=1}^{T} \left(\tau^{-(k+1)} \frac{\widehat{\theta} \sqrt{\epsilon_k}}{\sigma_{\mathbf{P}}^{\min}} \right) \left(\tau^{-(k-1)} \|\boldsymbol{\xi}_{k-1} - \boldsymbol{\xi}_\star\| \right) \\ &+ \sum_{k=1}^{T} \left(\tau^{-k} \frac{\bar{\theta}}{\sigma_{\mathbf{P}}^{\min}} \sqrt{\delta_k} \right) \left(\tau^{-k} \|\boldsymbol{\xi}_k - \boldsymbol{\xi}_\star\| \right). \end{aligned}$$

Applying Lemma A13 with

$$\beta_k \stackrel{\text{def}}{=} \tau^{-k} \|\boldsymbol{\xi}_k - \boldsymbol{\xi}_\star\|, \quad S_T \stackrel{\text{def}}{=} \frac{V_{\mathbf{P}}(\boldsymbol{\xi}_0)}{\sigma_{\mathbf{P}}^{\min}} + \frac{\widetilde{\theta}}{\sigma_{\mathbf{P}}^{\min}} \sum_{k=1}^{T} \tau^{-2k} \epsilon_k, \quad \widehat{\lambda}_k \stackrel{\text{def}}{=} \tau^{-(k+1)} \frac{\widehat{\theta} \sqrt{\epsilon_k}}{\sigma_{\mathbf{P}}^{\min}}, \quad \bar{\lambda}_k \stackrel{\text{def}}{=} \tau^{-k} \frac{\bar{\theta}}{\sigma_{\mathbf{P}}^{\min}} \sqrt{\delta_k},$$

we obtain

$$\begin{aligned} \tau^{-2T} \|\boldsymbol{\xi}_T - \boldsymbol{\xi}_\star\|^2 &\leq \left(\sqrt{ \frac{V_{\mathbf{P}}(\boldsymbol{\xi}_0)}{\sigma_{\mathbf{P}}^{\min}} + \frac{\widetilde{\theta}}{\sigma_{\mathbf{P}}^{\min}} \sum_{k=1}^{T} \tau^{-2k} \epsilon_k } + \sum_{k=1}^{T} \left(\tau^{-(k+1)} \frac{\widehat{\theta} \sqrt{\epsilon_k}}{\sigma_{\mathbf{P}}^{\min}} + \tau^{-k} \frac{\bar{\theta}}{\sigma_{\mathbf{P}}^{\min}} \sqrt{\delta_k} \right) \right)^2 \\ &\leq \left(\sqrt{\frac{V_{\mathbf{P}}(\boldsymbol{\xi}_0)}{\sigma_{\mathbf{P}}^{\min}}} + \sqrt{ \frac{\widetilde{\theta}}{\sigma_{\mathbf{P}}^{\min}} \sum_{k=1}^{T} \tau^{-2k} \epsilon_k } + \frac{\widehat{\theta}}{\tau \sigma_{\mathbf{P}}^{\min}} \sum_{k=1}^{T} \tau^{-k} \sqrt{\epsilon_k} + \frac{\bar{\theta}}{\sigma_{\mathbf{P}}^{\min}} \sum_{k=1}^{T} \tau^{-k} \sqrt{\delta_k} \right)^2 \\ &\leq \left(\sqrt{\frac{V_{\mathbf{P}}(\boldsymbol{\xi}_0)}{\sigma_{\mathbf{P}}^{\min}}} + \sqrt{ \frac{\widetilde{\theta}}{\sigma_{\mathbf{P}}^{\min}} } \sum_{k=1}^{T} \tau^{-k} \sqrt{\epsilon_k} + \frac{\widehat{\theta}}{\tau \sigma_{\mathbf{P}}^{\min}} \sum_{k=1}^{T} \tau^{-k} \sqrt{\epsilon_k} + \frac{\bar{\theta}}{\sigma_{\mathbf{P}}^{\min}} \sum_{k=1}^{T} \tau^{-k} \sqrt{\delta_k} \right)^2. \end{aligned}$$

It can be rewritten as

$$\|\boldsymbol{\xi}_T - \boldsymbol{\xi}_\star\|^2 \leq \tau^{2T} \left(\sqrt{\frac{V_{\mathbf{P}}(\boldsymbol{\xi}_0)}{\sigma_{\mathbf{P}}^{\min}}} + \left(\sqrt{\frac{\widetilde{\theta}}{\sigma_{\mathbf{P}}^{\min}}} + \frac{\widehat{\theta}}{\tau \sigma_{\mathbf{P}}^{\min}} \right) \sum_{k=1}^{T} \tau^{-k} \sqrt{\epsilon_k} + \frac{\bar{\theta}}{\sigma_{\mathbf{P}}^{\min}} \sum_{k=1}^{T} \tau^{-k} \sqrt{\delta_k} \right)^2.$$

Substituting $\boldsymbol{\varphi}$, it becomes

$$\|\boldsymbol{\varphi}_T - \boldsymbol{\varphi}_\star\| \le \tau^T \left(\sqrt{\kappa_{\mathbf{P}}}\|\boldsymbol{\varphi}_0 - \boldsymbol{\varphi}_\star\| + \left(\sqrt{\frac{\widetilde{\theta}}{\sigma_{\mathbf{P}}^{\min}}} + \frac{\widehat{\theta}}{\tau\sigma_{\mathbf{P}}^{\min}} \right) \sum_{k=1}^{T} \tau^{-k}\sqrt{\epsilon_k} + \frac{\bar{\theta}}{\sigma_{\mathbf{P}}^{\min}} \sum_{k=1}^{T} \tau^{-k}\sqrt{\delta_k} \right).$$

This completes the proof. ∎

A.4.1. Proof of Lemma A13

Proof This lemma is obtained by slightly modifying Lemma 1 presented in (**?**). We prove it by induction. It is true for $T = 0$ by assumption. Next, we assume it is true for $(T-1)$, then prove it is also true for T. We denote $\widetilde{\beta}_{T-1} \stackrel{\text{def}}{=} \max\{\beta_0, \dots, \beta_{T-1}\}$. It leads to

$$\beta_T^2 \le S_T + \widetilde{\beta}_{T-1}\left(\sum_{k=1}^{T}\widehat{\lambda}_k + \sum_{k=1}^{T-1}\overline{\lambda}_k\right) + \overline{\lambda}_T\beta_T \Rightarrow \left(\beta_T - \frac{\overline{\lambda}_T}{2}\right)^2 \le S_T + \frac{\overline{\lambda}_T^2}{4} + \widetilde{\beta}_{T-1}\left(\sum_{k=1}^{T}\widehat{\lambda}_k + \sum_{k=1}^{T-1}\overline{\lambda}_k\right)$$

The definition of $\widetilde{\beta}_T$ implies

$$\widetilde{\beta}_T = \max\{\beta_0, \dots, \beta_{T-1}, \beta_T\} \le \max\left\{ \widetilde{\beta}_{T-1}, \frac{\overline{\lambda}_T}{2} + \left(S_T + \frac{\overline{\lambda}_T^2}{4} + \widetilde{\beta}_{T-1}\left(\sum_{k=1}^{T}\widehat{\lambda}_k + \sum_{k=1}^{T-1}\overline{\lambda}_k\right)\right)^{1/2} \right\}.$$

The two terms in the maximum are equal if $\left(\widetilde{\beta}_{T-1} - \frac{\overline{\lambda}_T}{2}\right)^2 = S_T + \frac{\overline{\lambda}_T^2}{4} + \widetilde{\beta}_{T-1}\left(\sum_{k=1}^{T}\widehat{\lambda}_k + \sum_{k=1}^{T-1}\overline{\lambda}_k\right)$, i.e.,

$$\widetilde{\beta}_{T-1}^\star = \frac{1}{2}\sum_{k=1}^{T}\left(\widehat{\lambda}_k + \overline{\lambda}_k\right) + \left(S_T + \left(\frac{1}{2}\sum_{k=1}^{T}\left(\widehat{\lambda}_k + \overline{\lambda}_k\right)\right)^2\right)^{1/2}.$$

Then, we can consider two cases.
Case 1: If $\widetilde{\beta}_{T-1} \le \widetilde{\beta}_{T-1}^\star$, then $\widehat{\beta}_T \le \widehat{\beta}_{T-1}^\star$ since the two terms in the maximum are increasing function of $\widetilde{\beta}_{T-1}$.
Case 2: If $\widetilde{\beta}_{T-1} > \widetilde{\beta}_{T-1}^\star$, then

$$\widetilde{\beta}_{T-1} > \widetilde{\beta}_{T-1}^\star = \frac{1}{2}\sum_{k=1}^{T}\left(\widehat{\lambda}_k + \overline{\lambda}_k\right) + \left(S_T + \left(\frac{1}{2}\sum_{k=1}^{T}\left(\widehat{\lambda}_k + \overline{\lambda}_k\right)\right)^2\right)^{1/2}.$$

This leads to a contradiction as we assume (A37) holds for $(T-1)$. Combining two cases together, we obtain

$$\beta_T \le \widetilde{\beta}_T \le \frac{1}{2}\sum_{k=1}^{T}\left(\widehat{\lambda}_k + \overline{\lambda}_k\right) + \left(S_T + \left(\frac{1}{2}\sum_{k=1}^{T}\left(\widehat{\lambda}_k + \overline{\lambda}_k\right)\right)^2\right)^{1/2}.$$

Next, we prove the (A38). It can be proved by applying (A37). Relaxing the right-hand side of (A37), we obtain

$$\beta_k \leq \sqrt{S_k} + \sum_{i=1}^{k} \left(\widehat{\lambda}_i + \overline{\lambda}_i\right), \forall k \geq 0.$$

Thus,

$$\begin{aligned}
\beta_T^2 &\leq S_T + \sum_{k=1}^{T} \widehat{\lambda}_k \beta_{k-1} + \sum_{k=1}^{T} \overline{\lambda}_k \beta_k \\
&\leq S_T + \sum_{k=1}^{T} \widehat{\lambda}_k \left(\sqrt{S_{k-1}} + \sum_{i=1}^{k-1} \left(\widehat{\lambda}_i + \overline{\lambda}_i\right) \right) + \sum_{k=1}^{T} \overline{\lambda}_k \left(\sqrt{S_k} + \sum_{i=1}^{k} \left(\widehat{\lambda}_i + \overline{\lambda}_i\right) \right) \\
&\leq S_T + \left(\sqrt{S_T} + \sum_{k=1}^{T} \left(\widehat{\lambda}_k + \overline{\lambda}_k\right) \right) \sum_{k=1}^{T} \left(\widehat{\lambda}_k + \overline{\lambda}_k\right).
\end{aligned}$$

It implies

$$\beta_T^2 \leq \left(\sqrt{S_T} + \sum_{k=1}^{T} \left(\widehat{\lambda}_k + \overline{\lambda}_k\right) \right)^2.$$

This completes the proof. ■

Appendix B. Missing Proofs of Section 4

B.1. Proof of Lemma 8

Proof Since $\widehat{f}_\rho(\mathbf{w})$ is convex and $\widehat{L}/\rho$-smooth, and $\widehat{g}_\rho(\mathbf{w})$ is convex,

$$\begin{aligned}
\widehat{f}_\rho(\mathbf{w}_\star) &\geq \widehat{f}_\rho(\mathbf{w}_{k+1}) + \langle \boldsymbol{\beta}_{k+1}, \mathbf{w}_\star - \mathbf{w}_{k+1} \rangle + \frac{\rho}{2\widehat{L}} \|\boldsymbol{\beta}_{k+1} - \boldsymbol{\beta}_\star\|^2, \\
\widehat{g}_\rho(\mathbf{s}_\star) &\geq \widehat{g}_\rho(\mathbf{s}_{k+1}) + \langle \boldsymbol{\gamma}_{k+1}, \mathbf{s}_\star - \mathbf{s}_{k+1} \rangle.
\end{aligned}$$

Summing up them together,

$$\widehat{f}_\rho(\mathbf{w}_\star) + \widehat{g}_\rho(\mathbf{s}_\star) \geq \widehat{f}_\rho(\mathbf{w}_{k+1}) + \widehat{g}_\rho(\mathbf{s}_{k+1}) + \langle \boldsymbol{\beta}_{k+1}, \mathbf{w}_\star - \mathbf{w}_{k+1} \rangle + \langle \boldsymbol{\gamma}_{k+1}, \mathbf{s}_\star - \mathbf{s}_{k+1} \rangle + \frac{\rho}{2\widehat{L}} \|\boldsymbol{\beta}_{k+1} - \boldsymbol{\beta}_\star\|^2.$$

Rearranging both sides,

$$\begin{aligned}
&\widehat{f}_\rho(\mathbf{w}_\star) - \widehat{f}_\rho(\mathbf{w}_{k+1}) + \widehat{g}_\rho(\mathbf{s}_\star) - \widehat{g}_\rho(\mathbf{s}_{k+1}) - \langle \boldsymbol{\beta}_\star, \mathbf{w}_\star - \mathbf{w}_{k+1} \rangle - \langle \boldsymbol{\gamma}_\star, \mathbf{s}_\star - \mathbf{s}_{k+1} \rangle \\
\geq\ & - \langle \boldsymbol{\beta}_{k+1} - \boldsymbol{\beta}_\star, \mathbf{w}_{k+1} - \mathbf{w}_\star \rangle - \langle \boldsymbol{\gamma}_{k+1} - \boldsymbol{\gamma}_\star, \mathbf{s}_{k+1} - \mathbf{s}_\star \rangle + \frac{\rho}{2\widehat{L}} \|\boldsymbol{\beta}_{k+1} - \boldsymbol{\beta}_\star\|^2. \qquad \text{(A40)}
\end{aligned}$$

The optimality conditions of $(\mathbf{w}_\star, \mathbf{s}_\star, \mathbf{u}_\star)$ imply

$$\boldsymbol{\beta}_\star - \mathbf{u}_\star = \mathbf{0}, \quad \boldsymbol{\gamma}_\star - \mathbf{u}_\star = \mathbf{0}, \quad \mathbf{w}_\star + \mathbf{s}_\star - \mathbf{c} = \mathbf{0}.$$

Substituting these into (A40), we obtain

$$\widehat{f}_\rho(\mathbf{w}_\star) - \widehat{f}_\rho(\mathbf{w}_{k+1}) + \widehat{g}_\rho(\mathbf{s}_\star) - \widehat{g}_\rho(\mathbf{s}_{k+1}) - \langle \mathbf{u}_\star, \mathbf{w}_\star - \mathbf{w}_{k+1} \rangle - \langle \mathbf{u}_\star, \mathbf{s}_\star - \mathbf{s}_{k+1} \rangle + \langle \mathbf{w}_\star + \mathbf{s}_\star - \mathbf{c}, \mathbf{u}_\star - \mathbf{u}_{k+1} \rangle$$

$$\geq -\langle \boldsymbol{\beta}_{k+1}-\boldsymbol{\beta}_\star, \mathbf{w}_{k+1}-\mathbf{w}_\star\rangle - \langle \boldsymbol{\gamma}_{k+1}-\boldsymbol{\gamma}_\star, \mathbf{s}_{k+1}-\mathbf{s}_\star\rangle + \frac{\rho}{2\widehat{L}}\|\boldsymbol{\beta}_{k+1}-\boldsymbol{\beta}_\star\|^2.$$

It can be rewritten as

$$\begin{aligned}
&-\langle \boldsymbol{\beta}_{k+1}-\boldsymbol{\beta}_\star, \mathbf{w}_{k+1}-\mathbf{w}_\star\rangle - \langle \boldsymbol{\gamma}_{k+1}-\boldsymbol{\gamma}_\star, \mathbf{s}_{k+1}-\mathbf{s}_\star\rangle + \frac{\rho}{2\widehat{L}}\|\boldsymbol{\beta}_{k+1}-\boldsymbol{\beta}_\star\|^2 \\
&\leq \widehat{f}_\rho(\mathbf{w}_\star) - \widehat{f}_\rho(\mathbf{w}_{k+1}) + \widehat{g}_\rho(\mathbf{s}_\star) - \widehat{g}_\rho(\mathbf{s}_{k+1}) - \begin{bmatrix}\mathbf{w}_{k+1}-\mathbf{w}_\star\\ \mathbf{s}_{k+1}-\mathbf{s}_\star\\ \mathbf{u}_{k+1}-\mathbf{u}_\star\end{bmatrix}^\top \begin{bmatrix}-\mathbf{u}_\star\\ -\mathbf{u}_\star\\ \mathbf{w}_\star+\mathbf{s}_\star-\mathbf{c}\end{bmatrix}.
\end{aligned}$$

On the other hand, $S(\boldsymbol{\xi}_k, \mathbf{v}_k)$ can be written as

$$\begin{aligned}
S(\boldsymbol{\xi}_k, \mathbf{v}_k) &= \begin{bmatrix}\boldsymbol{\xi}_k-\boldsymbol{\xi}_\star\\ \mathbf{v}_k-\mathbf{v}_\star\end{bmatrix}^\top \left(\left(\begin{bmatrix}\widehat{\mathbf{C}}^1 & \widehat{\mathbf{D}}^1\\ \widehat{\mathbf{C}}^2 & \widehat{\mathbf{D}}^2\end{bmatrix}^\top \begin{bmatrix}\widehat{\mathbf{M}}^1 & \mathbf{0}\\ \mathbf{0} & \widehat{\mathbf{M}}^2\end{bmatrix} \begin{bmatrix}\widehat{\mathbf{C}}^1 & \widehat{\mathbf{D}}^1\\ \widehat{\mathbf{C}}^2 & \widehat{\mathbf{D}}^2\end{bmatrix}\right) \otimes \mathbf{I}_p\right) \begin{bmatrix}\boldsymbol{\xi}_k-\boldsymbol{\xi}_\star\\ \mathbf{v}_k-\mathbf{v}_\star\end{bmatrix}\\
&= \begin{bmatrix}\mathbf{y}_k^1-\mathbf{y}_\star^1\\ \mathbf{y}_k^2-\mathbf{y}_\star^2\end{bmatrix}^\top \left(\begin{bmatrix}\widehat{\mathbf{M}}^1 & \mathbf{0}\\ \mathbf{0} & \widehat{\mathbf{M}}^2\end{bmatrix} \otimes \mathbf{I}_p\right) \begin{bmatrix}\mathbf{y}_k^1-\mathbf{y}_\star^1\\ \mathbf{y}_k^2-\mathbf{y}_\star^2\end{bmatrix}.
\end{aligned}$$

Substituting $\widehat{\mathbf{M}}^1$ and $\widehat{\mathbf{M}}^2$, it becomes

$$\begin{aligned}
S(\boldsymbol{\xi}_k, \mathbf{v}_k) = &-\langle \boldsymbol{\beta}_{k+1}-\boldsymbol{\beta}_\star, \mathbf{w}_{k+1}-\mathbf{w}_\star\rangle - \langle \boldsymbol{\gamma}_{k+1}-\boldsymbol{\gamma}_\star, \mathbf{s}_{k+1}-\mathbf{s}_\star\rangle + \frac{\rho}{2\widehat{L}}\|\boldsymbol{\beta}_{k+1}-\boldsymbol{\beta}_\star\|^2\\
&-\langle \boldsymbol{\eta}_{k+1}, \mathbf{w}_{k+1}-\mathbf{w}_\star\rangle + \frac{\rho}{\widehat{L}}\langle \boldsymbol{\eta}_{k+1}, \boldsymbol{\beta}_{k+1}-\boldsymbol{\beta}_\star\rangle + \frac{\rho}{2\widehat{L}}\|\boldsymbol{\eta}_{k+1}\|^2 - \langle \mathbf{s}_{k+1}-\mathbf{s}_\star, \boldsymbol{\zeta}_{k+1}\rangle.
\end{aligned}$$

Applying (A40), we obtain

$$\begin{aligned}
S(\boldsymbol{\xi}_k, \mathbf{v}_k) \leq &\ \widehat{f}_\rho(\mathbf{w}_\star) - \widehat{f}_\rho(\mathbf{w}_{k+1}) + \widehat{g}_\rho(\mathbf{s}_\star) - \widehat{g}_\rho(\mathbf{s}_{k+1}) - \begin{bmatrix}\mathbf{w}_{k+1}-\mathbf{w}_\star\\ \mathbf{s}_{k+1}-\mathbf{s}_\star\\ \mathbf{u}_{k+1}-\mathbf{u}_\star\end{bmatrix}^\top \begin{bmatrix}-\mathbf{u}_\star\\ -\mathbf{u}_\star\\ \mathbf{w}_\star+\mathbf{s}_\star-\mathbf{c}\end{bmatrix}\\
&-\langle \boldsymbol{\eta}_{k+1}, \mathbf{w}_{k+1}-\mathbf{w}_\star\rangle + \frac{\rho}{\widehat{L}}\langle \boldsymbol{\eta}_{k+1}, \boldsymbol{\beta}_{k+1}-\boldsymbol{\beta}_\star\rangle + \frac{\rho}{2\widehat{L}}\|\boldsymbol{\eta}_{k+1}\|^2 - \langle \mathbf{s}_{k+1}-\mathbf{s}_\star, \boldsymbol{\zeta}_{k+1}\rangle.
\end{aligned}$$

Rearranging both sides, it becomes

$$\begin{aligned}
S(\boldsymbol{\xi}_k, \mathbf{v}_k) \leq &\widehat{f}_\rho(\mathbf{w}_\star) - \widehat{f}_\rho(\mathbf{w}_{k+1}) + \widehat{g}_\rho(\mathbf{s}_\star) - \widehat{g}_\rho(\mathbf{s}_{k+1}) - \begin{bmatrix}\mathbf{w}_{k+1}-\mathbf{w}_\star\\ \mathbf{s}_{k+1}-\mathbf{s}_\star\\ \mathbf{u}_{k+1}-\mathbf{u}_\star\end{bmatrix}^\top \begin{bmatrix}-\mathbf{u}_\star\\ -\mathbf{u}_\star\\ \mathbf{w}_\star+\mathbf{s}_\star-\mathbf{c}\end{bmatrix}\\
&+\left\langle \boldsymbol{\eta}_{k+1}, \frac{\rho}{\widehat{L}}(\boldsymbol{\beta}_{k+1}-\boldsymbol{\beta}_\star) - (\mathbf{w}_{k+1}-\mathbf{w}_\star)\right\rangle + \frac{\rho}{2\widehat{L}}\|\boldsymbol{\eta}_{k+1}\|^2 - \langle \mathbf{s}_{k+1}-\mathbf{s}_\star, \boldsymbol{\zeta}_{k+1}\rangle.
\end{aligned}$$

This completes the proof. ■

B.2. Proof of Theorem 9

Proof By the fact that $\boldsymbol{\xi}_\star$ is a fixed point of (14),

$$V_{\mathbf{P}}(\boldsymbol{\xi}_{k+1}) - V_{\mathbf{P}}(\boldsymbol{\xi}_k) = (\boldsymbol{\xi}_{k+1}-\boldsymbol{\xi}_\star)^\top \mathbf{P}(\boldsymbol{\xi}_{k+1}-\boldsymbol{\xi}_\star) - (\boldsymbol{\xi}_k-\boldsymbol{\xi}_\star)^\top \mathbf{P}(\boldsymbol{\xi}_k-\boldsymbol{\xi}_\star)$$

$$= \begin{bmatrix} \boldsymbol{\xi}_k - \boldsymbol{\xi}_\star \\ \mathbf{v}_k - \mathbf{v}_\star \end{bmatrix}^\top \left(\begin{bmatrix} \widehat{\mathbf{A}}^\top \widehat{\mathbf{P}} \widehat{\mathbf{A}} - \widehat{\mathbf{P}} & \widehat{\mathbf{A}}^\top \widehat{\mathbf{P}} \widehat{\mathbf{B}} \\ \widehat{\mathbf{B}}^\top \widehat{\mathbf{P}} \widehat{\mathbf{A}} & \widehat{\mathbf{B}}^\top \widehat{\mathbf{P}} \widehat{\mathbf{B}} \end{bmatrix} \otimes \mathbf{I}_p \right) \begin{bmatrix} \boldsymbol{\xi}_k - \boldsymbol{\xi}_\star \\ \mathbf{v}_k - \mathbf{v}_\star \end{bmatrix}.$$

Applying (27), it becomes

$$V_{\mathbf{P}}(\boldsymbol{\xi}_{k+1}) - V_{\mathbf{P}}(\boldsymbol{\xi}_k) \le S(\boldsymbol{\xi}_k, \mathbf{v}_k).$$

Substituting the upper bound of $S(\boldsymbol{\xi}_k, \mathbf{v}_k)$ from Lemma 8, we obtain

$$\begin{aligned} &\widehat{f}_\rho(\mathbf{w}_{k+1}) - \widehat{f}_\rho(\mathbf{w}_\star) + \widehat{g}_\rho(\mathbf{s}_{k+1}) - \widehat{g}_\rho(\mathbf{s}_\star) + \begin{bmatrix} \mathbf{w}_{k+1} - \mathbf{w}_\star \\ \mathbf{s}_{k+1} - \mathbf{s}_\star \\ \mathbf{u}_{k+1} - \mathbf{u}_\star \end{bmatrix}^\top \begin{bmatrix} -\mathbf{u}_\star \\ -\mathbf{u}_\star \\ \mathbf{w}_\star + \mathbf{s}_\star - \mathbf{c} \end{bmatrix} + V_{\mathbf{P}}(\boldsymbol{\xi}_{k+1}) \\ &\le V_{\mathbf{P}}(\boldsymbol{\xi}_k) + \left\langle \boldsymbol{\eta}_{k+1}, \frac{\rho}{\widehat{L}}(\boldsymbol{\beta}_{k+1} - \boldsymbol{\beta}_\star) - (\mathbf{w}_{k+1} - \mathbf{w}_\star) \right\rangle + \frac{\rho}{2\widehat{L}} \|\boldsymbol{\eta}_{k+1}\|^2 - \langle \mathbf{s}_{k+1} - \mathbf{s}_\star, \boldsymbol{\zeta}_{k+1} \rangle \\ &\le V_{\mathbf{P}}(\boldsymbol{\xi}_k) + \left\| \frac{\rho}{\widehat{L}}(\boldsymbol{\beta}_{k+1} - \boldsymbol{\beta}_\star) - (\mathbf{w}_{k+1} - \mathbf{w}_\star) \right\| \|\boldsymbol{\eta}_{k+1}\| + \frac{\rho}{2\widehat{L}} \|\boldsymbol{\eta}_{k+1}\|^2 + \|\mathbf{s}_{k+1} - \mathbf{s}_\star\| \|\boldsymbol{\zeta}_{k+1}\| \\ &\le V_{\mathbf{P}}(\boldsymbol{\xi}_k) + \frac{\rho + \widehat{L}}{\widehat{L}} \big(\|\boldsymbol{\xi}_k - \boldsymbol{\xi}_\star\| + \|\boldsymbol{\eta}_{k+1}\| \big) \|\boldsymbol{\eta}_{k+1}\| + \frac{\rho}{2\widehat{L}} \|\boldsymbol{\eta}_{k+1}\|^2 + \|\boldsymbol{\xi}_{k+1} - \boldsymbol{\xi}_\star\| \|\boldsymbol{\zeta}_{k+1}\| \\ &= V_{\mathbf{P}}(\boldsymbol{\xi}_k) + \frac{\rho + \widehat{L}}{\widehat{L}} \|\boldsymbol{\xi}_k - \boldsymbol{\xi}_\star\| \|\boldsymbol{\eta}_{k+1}\| + \frac{3\rho + 2\widehat{L}}{2\widehat{L}} \|\boldsymbol{\eta}_{k+1}\|^2 + \|\boldsymbol{\xi}_{k+1} - \boldsymbol{\xi}_\star\| \|\boldsymbol{\zeta}_{k+1}\|, \end{aligned}$$

where the last inequality follows from Lemma A12. Substituting $\|\boldsymbol{\eta}_{k+1}\|^2 \le 2\epsilon_{k+1}/\rho$ and $\|\boldsymbol{\zeta}_{k+1}\|^2 \le 2\delta_{k+1}/\rho$,

$$\begin{aligned} &\widehat{f}_\rho(\mathbf{w}_{k+1}) - \widehat{f}_\rho(\mathbf{w}_\star) + \widehat{g}_\rho(\mathbf{s}_{k+1}) - \widehat{g}_\rho(\mathbf{s}_\star) + \begin{bmatrix} \mathbf{w}_{k+1} - \mathbf{w}_\star \\ \mathbf{s}_{k+1} - \mathbf{s}_\star \\ \mathbf{u}_{k+1} - \mathbf{u}_\star \end{bmatrix}^\top \begin{bmatrix} -\mathbf{u}_\star \\ -\mathbf{u}_\star \\ \mathbf{w}_\star + \mathbf{s}_\star - \mathbf{c} \end{bmatrix} + V_{\mathbf{P}}(\boldsymbol{\xi}_{k+1}) \\ &\le V_{\mathbf{P}}(\boldsymbol{\xi}_k) + \frac{\rho + \widehat{L}}{\widehat{L}} \sqrt{\frac{2}{\rho}} \|\boldsymbol{\xi}_k - \boldsymbol{\xi}_\star\| \sqrt{\epsilon_{k+1}} + \frac{3\rho + 2\widehat{L}}{\rho \widehat{L}} \epsilon_{k+1} + \sqrt{\frac{2}{\rho}} \|\boldsymbol{\xi}_{k+1} - \boldsymbol{\xi}_\star\| \sqrt{\delta_{k+1}}. \end{aligned}$$

Summing up it from $k = 0$ to $k = T - 1$,

$$\begin{aligned} &\sum_{k=1}^{T} \left\{ \widehat{f}_\rho(\mathbf{w}_k) - \widehat{f}_\rho(\mathbf{w}_\star) + \widehat{g}_\rho(\mathbf{s}_k) - \widehat{g}_\rho(\mathbf{s}_\star) + \begin{bmatrix} \mathbf{w}_k - \mathbf{w}_\star \\ \mathbf{s}_k - \mathbf{s}_\star \\ \mathbf{u}_k - \mathbf{u}_\star \end{bmatrix}^\top \begin{bmatrix} -\mathbf{u}_\star \\ -\mathbf{u}_\star \\ \mathbf{w}_\star + \mathbf{s}_\star - \mathbf{c} \end{bmatrix} \right\} + V_{\mathbf{P}}(\boldsymbol{\xi}_T) \\ &\le V_{\mathbf{P}}(\boldsymbol{\xi}_0) + \frac{3\rho + 2\widehat{L}}{\rho \widehat{L}} \sum_{k=1}^{T} \epsilon_k + \frac{\rho + \widehat{L}}{\widehat{L}} \sqrt{\frac{2}{\rho}} \sum_{k=1}^{T} \|\boldsymbol{\xi}_{k-1} - \boldsymbol{\xi}_\star\| \sqrt{\epsilon_k} + \sqrt{\frac{2}{\rho}} \sum_{k=1}^{T} \|\boldsymbol{\xi}_k - \boldsymbol{\xi}_\star\| \sqrt{\delta_k}. \end{aligned}$$

By the definitions of $\mathbf{w}, \mathbf{s}, \widehat{f}_\rho(\mathbf{w})$ and $\widehat{g}_\rho(\mathbf{s})$, it can be rewritten as

$$\frac{1}{\rho} \sum_{k=1}^{T} \left\{ f(\mathbf{x}_k) - f(\mathbf{x}_\star) + g(\mathbf{z}_k) - g(\mathbf{z}_\star) + \begin{bmatrix} \mathbf{x}_k - \mathbf{x}_\star \\ \mathbf{z}_k - \mathbf{z}_\star \\ \rho(\mathbf{u}_k - \mathbf{u}_\star) \end{bmatrix}^\top \begin{bmatrix} -\rho \mathbf{A}^\top \mathbf{u}_\star \\ -\rho \mathbf{B}^\top \mathbf{u}_\star \\ \mathbf{A}\mathbf{x}_\star + \mathbf{B}\mathbf{z}_\star - \mathbf{c} \end{bmatrix} \right\} + V_{\mathbf{P}}(\boldsymbol{\xi}_T)$$

$$\leq V_{\mathbf{P}}(\boldsymbol{\xi}_0) + \frac{3\rho + 2\widehat{L}}{\rho\widehat{L}} \sum_{k=1}^{T} \epsilon_k + \frac{\rho + \widehat{L}}{\widehat{L}} \sqrt{\frac{2}{\rho}} \sum_{k=1}^{T} \|\boldsymbol{\xi}_{k-1} - \boldsymbol{\xi}_\star\| \sqrt{\epsilon_k} + \sqrt{\frac{2}{\rho}} \sum_{k=1}^{T} \|\boldsymbol{\xi}_k - \boldsymbol{\xi}_\star\| \sqrt{\delta_k}. \tag{A41}$$

Applying (23) to the right-hand side,

$$V_{\mathbf{P}}(\boldsymbol{\xi}_T) \leq V_{\mathbf{P}}(\boldsymbol{\xi}_0) + \frac{3\rho + 2\widehat{L}}{\rho\widehat{L}} \sum_{k=1}^{T} \epsilon_k + \frac{\rho + \widehat{L}}{\widehat{L}} \sqrt{\frac{2}{\rho}} \sum_{k=1}^{T} \|\boldsymbol{\xi}_{k-1} - \boldsymbol{\xi}_\star\| \sqrt{\epsilon_k} + \sqrt{\frac{2}{\rho}} \sum_{k=1}^{T} \|\boldsymbol{\xi}_k - \boldsymbol{\xi}_\star\| \sqrt{\delta_k}.$$

By further relaxing the right-hand side,

$$\|\boldsymbol{\xi}_T - \boldsymbol{\xi}_0\|^2 \leq \frac{V_{\mathbf{P}}(\boldsymbol{\xi}_0)}{\sigma_{\mathbf{P}}^{\min}} + \frac{3\rho + 2\widehat{L}}{\rho\widehat{L}\sigma_{\mathbf{P}}^{\min}} \sum_{k=1}^{T} \epsilon_k + \frac{\rho + \widehat{L}}{\widehat{L}\sigma_{\mathbf{P}}^{\min}} \sqrt{\frac{2}{\rho}} \sum_{k=1}^{T} \|\boldsymbol{\xi}_{k-1} - \boldsymbol{\xi}_\star\| \sqrt{\epsilon_k} + \frac{1}{\sigma_{\mathbf{P}}^{\min}} \sqrt{\frac{2}{\rho}} \sum_{k=1}^{T} \|\boldsymbol{\xi}_k - \boldsymbol{\xi}_\star\| \sqrt{\delta_k}.$$

Applying Lemma A13 with

$$\beta_k \stackrel{\text{def}}{=} \|\boldsymbol{\xi}_k - \boldsymbol{\xi}_0\|, \quad S_T \stackrel{\text{def}}{=} \frac{V_{\mathbf{P}}(\boldsymbol{\xi}_0)}{\sigma_{\mathbf{P}}^{\min}} + \frac{3\rho + 2\widehat{L}}{\rho\widehat{L}\sigma_{\mathbf{P}}^{\min}} \sum_{k=1}^{T} \epsilon_k, \quad \widehat{\lambda}_k \stackrel{\text{def}}{=} \frac{\rho + \widehat{L}}{\widehat{L}\sigma_{\mathbf{P}}^{\min}} \sqrt{\frac{2}{\rho}} \sqrt{\epsilon_k}, \quad \overline{\lambda}_k \stackrel{\text{def}}{=} \frac{1}{\sigma_{\mathbf{P}}^{\min}} \sqrt{\frac{2}{\rho}} \sqrt{\delta_k},$$

we obtain

$$\begin{aligned} &\frac{V_{\mathbf{P}}(\boldsymbol{\xi}_0)}{\sigma_{\mathbf{P}}^{\min}} + \frac{3\rho + 2\widehat{L}}{\rho\widehat{L}\sigma_{\mathbf{P}}^{\min}} \sum_{k=1}^{T} \epsilon_k + \frac{\rho + \widehat{L}}{\widehat{L}\sigma_{\mathbf{P}}^{\min}} \sqrt{\frac{2}{\rho}} \sum_{k=1}^{T} \|\boldsymbol{\xi}_{k-1} - \boldsymbol{\xi}_\star\| \sqrt{\epsilon_k} + \frac{1}{\sigma_{\mathbf{P}}^{\min}} \sqrt{\frac{2}{\rho}} \sum_{k=1}^{T} \|\boldsymbol{\xi}_k - \boldsymbol{\xi}_\star\| \sqrt{\delta_k} \\ &\leq \left(\sqrt{\frac{V_{\mathbf{P}}(\boldsymbol{\xi}_0)}{\sigma_{\mathbf{P}}^{\min}} + \frac{3\rho + 2\widehat{L}}{\rho\widehat{L}\sigma_{\mathbf{P}}^{\min}} \sum_{k=1}^{T} \epsilon_k} + \frac{\rho + \widehat{L}}{\widehat{L}\sigma_{\mathbf{P}}^{\min}} \sqrt{\frac{2}{\rho}} \sum_{k=1}^{T} \sqrt{\epsilon_k} + \frac{1}{\sigma_{\mathbf{P}}^{\min}} \sqrt{\frac{2}{\rho}} \sum_{k=1}^{T} \sqrt{\delta_k} \right)^2. \end{aligned}$$

It is equivalent to

$$\begin{aligned} &V_{\mathbf{P}}(\boldsymbol{\xi}_0) + \frac{3\rho + 2\widehat{L}}{\rho\widehat{L}} \sum_{k=1}^{T} \epsilon_k + \frac{\rho + \widehat{L}}{\widehat{L}} \sqrt{\frac{2}{\rho}} \sum_{k=1}^{T} \|\boldsymbol{\xi}_{k-1} - \boldsymbol{\xi}_\star\| \sqrt{\epsilon_k} + \sqrt{\frac{2}{\rho}} \sum_{k=1}^{T} \|\boldsymbol{\xi}_k - \boldsymbol{\xi}_\star\| \sqrt{\delta_k} \\ &\leq \left(\sqrt{V_{\mathbf{P}}(\boldsymbol{\xi}_0) + \frac{3\rho + 2\widehat{L}}{\rho\widehat{L}} \sum_{k=1}^{T} \epsilon_k} + \frac{\rho + \widehat{L}}{\widehat{L}} \sqrt{\frac{2}{\rho\sigma_{\mathbf{P}}^{\min}}} \sum_{k=1}^{T} \sqrt{\epsilon_k} + \sqrt{\frac{2}{\rho\sigma_{\mathbf{P}}^{\min}}} \sum_{k=1}^{T} \sqrt{\delta_k} \right)^2. \end{aligned}$$

By further relaxing right-hand side,

$$\begin{aligned} &V_{\mathbf{P}}(\boldsymbol{\xi}_0) + \frac{3\rho + 2\widehat{L}}{\rho\widehat{L}} \sum_{k=1}^{T} \epsilon_k + \frac{\rho + \widehat{L}}{\widehat{L}} \sqrt{\frac{2}{\rho}} \sum_{k=1}^{T} \|\boldsymbol{\xi}_{k-1} - \boldsymbol{\xi}_\star\| \sqrt{\epsilon_k} + \sqrt{\frac{2}{\rho}} \sum_{k=1}^{T} \|\boldsymbol{\xi}_k - \boldsymbol{\xi}_\star\| \sqrt{\delta_k} \\ &\leq \left(\sqrt{V_{\mathbf{P}}(\boldsymbol{\xi}_0)} + \left(\sqrt{\frac{3\rho + 2\widehat{L}}{\rho\widehat{L}}} + \frac{\rho + \widehat{L}}{\widehat{L}} \sqrt{\frac{2}{\rho\sigma_{\mathbf{P}}^{\min}}} \right) \sum_{k=1}^{T} \sqrt{\epsilon_k} + \sqrt{\frac{2}{\rho\sigma_{\mathbf{P}}^{\min}}} \sum_{k=1}^{T} \sqrt{\delta_k} \right)^2. \end{aligned}$$

Substituting it into (A41), we obtain

$$\frac{1}{\rho}\sum_{k=1}^{T}\left\{f(\mathbf{x}_k)-f(\mathbf{x}_\star)+g(\mathbf{z}_k)-g(\mathbf{z}_\star)+\begin{bmatrix}\mathbf{x}_k-\mathbf{x}_\star\\ \mathbf{z}_k-\mathbf{z}_\star\\ \rho(\mathbf{u}_k-\mathbf{u}_\star)\end{bmatrix}^\top\begin{bmatrix}-\rho\mathbf{A}^\top\mathbf{u}_\star\\ -\rho\mathbf{B}^\top\mathbf{u}_\star\\ \mathbf{A}\mathbf{x}_\star+\mathbf{B}\mathbf{z}_\star-\mathbf{c}\end{bmatrix}\right\}+V_{\mathbf{P}}(\boldsymbol{\xi}_k)$$
$$\leq\left(\sqrt{V_{\mathbf{P}}(\boldsymbol{\xi}_0)}+\left(\sqrt{\frac{3\rho+2\widehat{L}}{\rho\widehat{L}}}+\frac{\rho+\widehat{L}}{\widehat{L}}\sqrt{\frac{2}{\rho\sigma_{\mathbf{P}}^{\min}}}\right)\sum_{k=1}^{T}\sqrt{\epsilon_k}+\sqrt{\frac{2}{\rho\sigma_{\mathbf{P}}^{\min}}}\sum_{k=1}^{T}\sqrt{\delta_k}\right)^2.$$

Dividing both sides by T,

$$\frac{1}{T}\sum_{k=1}^{T}\left\{f(\mathbf{x}_k)-f(\mathbf{x}_\star)+g(\mathbf{z}_k)-g(\mathbf{z}_\star)+\begin{bmatrix}\mathbf{x}_k-\mathbf{x}_\star\\ \mathbf{z}_k-\mathbf{z}_\star\\ \rho(\mathbf{u}_k-\mathbf{u}_\star)\end{bmatrix}^\top\begin{bmatrix}-\rho\mathbf{A}^\top\mathbf{u}_\star\\ -\rho\mathbf{B}^\top\mathbf{u}_\star\\ \mathbf{A}\mathbf{x}_\star+\mathbf{B}\mathbf{z}_\star-\mathbf{c}\end{bmatrix}\right\}+\frac{\rho}{T}V_{\mathbf{P}}(\boldsymbol{\xi}_k)$$
$$\leq\frac{\rho}{T}\left(\sqrt{V_{\mathbf{P}}(\boldsymbol{\xi}_0)}+\left(\sqrt{\frac{3\rho+2\widehat{L}}{\rho\widehat{L}}}+\frac{\rho+\widehat{L}}{\widehat{L}}\sqrt{\frac{2}{\rho\sigma_{\mathbf{P}}^{\min}}}\right)\sum_{k=1}^{T}\sqrt{\epsilon_k}+\sqrt{\frac{2}{\rho\sigma_{\mathbf{P}}^{\min}}}\sum_{k=1}^{T}\sqrt{\delta_k}\right)^2.$$

Applying the convexity of $f(\mathbf{x})$ and $g(\mathbf{z})$,

$$f(\bar{\mathbf{x}}_T)-f(\mathbf{x}_\star)+g(\bar{\mathbf{z}}_T)-g(\mathbf{z}_\star)+\begin{bmatrix}\bar{\mathbf{x}}_T-\mathbf{x}_\star\\ \bar{\mathbf{z}}_T-\mathbf{z}_\star\\ \rho(\bar{\mathbf{u}}_T-\mathbf{u}_\star)\end{bmatrix}^\top\begin{bmatrix}-\rho\mathbf{A}^\top\mathbf{u}_\star\\ -\rho\mathbf{B}^\top\mathbf{u}_\star\\ \mathbf{A}\mathbf{x}_\star+\mathbf{B}\mathbf{z}_\star-\mathbf{c}\end{bmatrix}+\frac{\rho}{T}V_{\mathbf{P}}(\boldsymbol{\xi}_T)$$
$$\leq\frac{\rho}{T}\left(\sqrt{V_{\mathbf{P}}(\boldsymbol{\xi}_0)}+\left(\sqrt{\frac{3\rho+2\widehat{L}}{\rho\widehat{L}}}+\frac{\rho+\widehat{L}}{\widehat{L}}\sqrt{\frac{2}{\rho\sigma_{\mathbf{P}}^{\min}}}\right)\sum_{k=1}^{T}\sqrt{\epsilon_k}+\sqrt{\frac{2}{\rho\sigma_{\mathbf{P}}^{\min}}}\sum_{k=1}^{T}\sqrt{\delta_k}\right)^2.$$

Note that $V_{\mathbf{P}}(\boldsymbol{\xi}_T)\geq 0$, thus

$$f(\bar{\mathbf{x}}_T)-f(\mathbf{x}_\star)+g(\bar{\mathbf{z}}_T)-g(\mathbf{z}_\star)+\begin{bmatrix}\bar{\mathbf{x}}_T-\mathbf{x}_\star\\ \bar{\mathbf{z}}_T-\mathbf{z}_\star\\ \rho(\bar{\mathbf{u}}_T-\mathbf{u}_\star)\end{bmatrix}^\top\begin{bmatrix}-\rho\mathbf{A}^\top\mathbf{u}_\star\\ -\rho\mathbf{B}^\top\mathbf{u}_\star\\ \mathbf{A}\mathbf{x}_\star+\mathbf{B}\mathbf{z}_\star-\mathbf{c}\end{bmatrix}$$
$$\leq\frac{\rho}{T}\left(\sqrt{V_{\mathbf{P}}(\boldsymbol{\xi}_0)}+\left(\sqrt{\frac{3\rho+2\widehat{L}}{\rho\widehat{L}}}+\frac{\rho+\widehat{L}}{\widehat{L}}\sqrt{\frac{2}{\rho\sigma_{\mathbf{P}}^{\min}}}\right)\sum_{k=1}^{T}\sqrt{\epsilon_k}+\sqrt{\frac{2}{\rho\sigma_{\mathbf{P}}^{\min}}}\sum_{k=1}^{T}\sqrt{\delta_k}\right)^2.$$

This completes the proof. ■

B.3. Proof of Theorem 11

Proof Without loss of generality, we assume $\rho=2\rho_0\widehat{L}$ and $\widehat{\mathbf{P}}=x\mathbf{I}_2$ where $\rho_0,x>0$. Then, $\widehat{\mathbf{M}}^1$ becomes

$$\widehat{\mathbf{M}}^1=\begin{bmatrix}0 & -0.5\\ -0.5 & \rho_0\end{bmatrix}.$$

We define $\mathbf{S}$ as

$$\mathbf{S} = \begin{bmatrix} \widehat{\mathbf{A}}^\top \widehat{\mathbf{P}} \widehat{\mathbf{A}} - \widehat{\mathbf{P}} & \widehat{\mathbf{A}}^\top \widehat{\mathbf{P}} \widehat{\mathbf{B}} \\ \widehat{\mathbf{B}}^\top \widehat{\mathbf{P}} \widehat{\mathbf{A}} & \widehat{\mathbf{B}}^\top \widehat{\mathbf{P}} \widehat{\mathbf{B}} \end{bmatrix} - \begin{bmatrix} \widehat{\mathbf{C}}^1 & \widehat{\mathbf{D}}^1 \\ \widehat{\mathbf{C}}^2 & \widehat{\mathbf{D}}^2 \end{bmatrix}^\top \begin{bmatrix} \widehat{\mathbf{M}}^1 & \mathbf{0} \\ \mathbf{0} & \widehat{\mathbf{M}}^2 \end{bmatrix}.$$

Next, we show that $\det(\mathbf{S}) \leq 0$. Substituting the values of all matrices into $\mathbf{S}$, we obtain

$$\mathbf{S} \overset{\text{def}}{=} \begin{bmatrix} 0 & x(1-\alpha) & x\alpha - 0.5 & x - 0.5 \\ x(1-\alpha) & x\alpha(\alpha-2) & x\alpha(1-\alpha)+0.5 & (\alpha-1)(x-0.5) \\ x\alpha - 0.5 & x\alpha(1-\alpha)+0.5 & x\alpha^2 - (\rho_0+1) & -\alpha(x-0.5) \\ x-0.5 & (\alpha-1)(x-0.5) & -\alpha(x-0.5) & 2x-1 \end{bmatrix}.$$

Substituting $x = \frac{1}{2}$, $\mathbf{S}$ becomes

$$\mathbf{S} = \begin{bmatrix} 0 & 0.5(1-\alpha) & 0.5(\alpha-1) & 0 \\ 0.5(1-\alpha) & 0.5\alpha(\alpha-2) & 0.5\alpha(1-\alpha)+0.5 & 0 \\ 0.5(\alpha-1) & 0.5\alpha(1-\alpha)+0.5 & 0.5\alpha^2 - (\rho_0+1) & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

It is easy to see that $\det(\mathbf{S}) = 0$. Thus, $\widehat{\mathbf{P}} = 0.5\mathbf{I}_2$ satisfies the linear matrix inequality (27). This complete the proof. ■

Appendix C. More Details of Experiments

C.1. Problem Formulation and ADMM Updates

The distributed ℓ_1-norm regularized logistic regression can be written as

$$\begin{aligned} \min_{\{\mathbf{x}^j, v^j\}_{j=1}^J, \mathbf{z}} \quad & \sum_{j=1}^{J} \sum_{i=1}^{n_j} \log\Big(1 + \exp\Big(-b_i^j \Big(\langle \mathbf{a}_i^j, \mathbf{x}^j \rangle + v^j\Big)\Big)\Big) + \lambda \|\mathbf{z}\|_1 \\ \text{s.t. } & \mathbf{x}^j - \mathbf{z} = \mathbf{0}, \quad v^j = \eta, \quad \forall j = 1, \ldots, J. \end{aligned} \tag{A42}$$

The augmented Lagrangian of (A42) is formed as

$$\begin{aligned} L_\rho\big(\{\mathbf{x}^j, v^j\}_{j=1}^J, \mathbf{z}, \{\mathbf{u}^j, s^j\}_{j=1}^J\big) = & \sum_{j=1}^{J} \sum_{i=1}^{n_j} \log\Big(1 + \exp\Big(-b_i^j \Big(\langle \mathbf{a}_i^j, \mathbf{x}^j \rangle + v^j\Big)\Big)\Big) + \lambda \|\mathbf{z}\|_1 \\ & - \rho \sum_{j=1}^{J} \langle \mathbf{u}^j, \mathbf{x}^j - \mathbf{z} \rangle + \frac{\rho}{2} \sum_{j=1}^{J} \|\mathbf{x}^j - \mathbf{z}\|^2 - \rho \sum_{j=1}^{J} s^j(v^j - \eta) + \frac{\rho}{2} \sum_{j=1}^{J} (v^j - \eta)^2. \end{aligned}$$

Updates for $(\mathbf{x}^j, v^j)$
For $\forall j = 1, \ldots, J$, $\mathbf{x}_{k+1}^j$ and v_{k+1}^j can be obtained by solving

$$\min_{\mathbf{x}^j, v^j} \left\{ \sum_{i=1}^{n_j} \log\Big(1 + \exp\Big(-b_i^j \Big(\langle \mathbf{a}_i^j, \mathbf{x}^j \rangle + v^j\Big)\Big)\Big) + \frac{\rho}{2} \big\|\mathbf{x}^j - \mathbf{u}_k^j - \mathbf{z}_k\big\|^2 + \frac{\rho}{2} \big(v^j - s_k^j - \eta_k\big)^2 \right\}. \tag{A43}$$

It does not have an analytical solution, thus we apply L-BFGS to solve it. We need to compute the duality gap when it be used as the criteria to terminate L-BFGS solver. Thus, we derive the dual problem. For convenience, define $\overline{\mathbf{x}}_k^j$ and $\bar{v}_k^j$ as $\overline{\mathbf{x}}_k^j = \mathbf{u}_k^j + \mathbf{z}_k^j$ and $\bar{v}_k^j = s_k^j + \eta_k^j$. Then, (A43) can be written as

$$\min_{\mathbf{x}^j, v^j} \left\{ \sum_{i=1}^{n_j} \log\Big(1+\exp\Big(-b_i^j\Big(\langle \mathbf{a}_i^j, \mathbf{x}^j\rangle + v^j\Big)\Big)\Big) + \frac{\rho}{2}\big\|\mathbf{x}^j - \overline{\mathbf{x}}_k^j\big\|^2 + \frac{\rho}{2}(v^j - \bar{v}_k^j)^2 \right\}. \tag{A44}$$

The dual problem of (A44) is

$$\begin{aligned} \max_{\boldsymbol{\theta}^j} \Bigg\{ & -\sum_{i=1}^{n_j}\Big\{\theta_i^j \log\theta_i^j + (1-\theta_i^j)\log(1-\theta_i^j)\Big\} - \frac{1}{2\rho}\|\bar{\mathbf{A}}^j\boldsymbol{\theta}^j\|^2 - \langle \overline{\mathbf{x}}_k^j, \bar{\mathbf{A}}^j\boldsymbol{\theta}^j\rangle - \frac{1}{2\rho}\langle \mathbf{b}^j, \boldsymbol{\theta}^j\rangle^2 \\ & - \bar{v}_k^j \langle \mathbf{b}^j, \boldsymbol{\theta}^j\rangle \Bigg\} \\ \text{s.t. } & \theta_i^j \in (0,1), \end{aligned} \tag{A45}$$

where $\bar{\mathbf{A}}^j = [b_1^j\mathbf{a}_1^j, \ldots, b_{n_j}^j\mathbf{a}_{n_j}^j]$ and $\mathbf{b}^j = [b_1^j, \ldots, b_{n_j}^j]$. In addition, the KKT condition of (A44) establishes

$$(\theta_i^j)^\star = \frac{\exp\big(-b_i^j\big(\langle \mathbf{a}_i^j, (\mathbf{x}^j)^\star\rangle + (v^j)^\star\big)\big)}{1+\exp\big(-b_i^j\big(\langle \mathbf{a}_i^j, (\mathbf{x}^j)^\star\rangle + (v^j)^\star\big)\big)}. \tag{A46}$$

Thus, given a primal solution $\mathbf{x}^j, v^j$, we can get a dual solution θ^j as

$$\theta_i^j = \frac{\exp\big(-b_i^j\big(\langle \mathbf{a}_i^j, \mathbf{x}^j\rangle + v^j\big)\big)}{1+\exp\big(-b_i^j\big(\langle \mathbf{a}_i^j, \mathbf{x}^j\rangle + v^j\big)\big)}, \forall i = 1, \ldots n_j. \tag{A47}$$

It is straightforward to show $\boldsymbol{\theta}^j$ is a feasible solution. Then, the dual gap $G(\mathbf{x}^j, \boldsymbol{\theta}^j)$ can be computed by applying (A43) and (A45).

Updates for $(\mathbf{Z}, \eta)$

Specifically, $\mathbf{z}$ and η can be obtained by solving

$$\min_{\mathbf{z},\eta}\left\{\lambda\|\mathbf{z}\|_1 + \sum_{j=1}^{J}\frac{\rho}{2}\big\|\alpha\mathbf{x}_{k+1}^j + (1-\alpha)\mathbf{z}_k - \mathbf{u}_k^j - \mathbf{z}\big\|^2 + \sum_{j=1}^{J}\frac{\rho}{2}\big(\alpha v_{k+1}^j + (1-\alpha)\eta_k - s_k^j - \eta\big)^2\right\}.$$

Thus, we obtain

$$\mathbf{z}_{k+1} = \mathcal{S}_{\frac{\lambda}{\rho J}}\left(\frac{1}{J}\sum_{j=1}^{J}\Big(\alpha\mathbf{x}_{k+1}^j + (1-\alpha)\mathbf{z}_k - \mathbf{u}_k^j\Big)\right) \text{ and } \eta_{k+1} = \frac{1}{J}\sum_{j=1}^{J}\Big(\alpha v_{k+1}^j + (1-\alpha)\eta_k - s_k^j\Big),$$

where $\mathcal{S}_{\frac{\lambda}{\rho J}}(\cdot)$ denotes the Soft Thresholding operator for ℓ_1-norm.

Updates for $(\mathbf{u}^j, s^j)$

For $\forall j = 1, \ldots, J$, $\mathbf{u}^j$ and s^j can be updated as following

$$\mathbf{u}_{k+1}^j = \mathbf{u}_k^j - \big(\alpha\mathbf{x}_{k+1}^j + (1-\alpha)\mathbf{z}_k - \mathbf{z}_{k+1}\big) \text{ and } s_{k+1}^j = s_k^j - \big(\alpha v_{k+1}^j + (1-\alpha)\eta_k - \eta_{k+1}\big).$$

C.2. Experiment Setting

We define $n = \sum_{j=1}^{J} n_j$ as the total number of samples over all workers. We use n^+ and n^- to denote the total number of positive and negative samples, respectively, overall all workers. By collecting the data over all workers, (A42) is equivalent to

$$\min_{\mathbf{x},v,\mathbf{z}} \sum_{i=1}^{n} \log(1+\exp(-b_i(\langle \mathbf{a}_i, \mathbf{x}\rangle + v))) + \lambda\|\mathbf{z}\|_1 \quad \text{s.t.} \quad \mathbf{x}-\mathbf{z}=\mathbf{0}. \tag{A48}$$

Koh et al. (2007) show that the dual problem of (A48) is

$$\max_{\boldsymbol{\beta}} \quad -\sum_{i=1}^{n} \Big\{\beta_i \log\beta_i + (1-\beta_i)\log(1-\beta_i)\Big\} \;\text{ s.t. }\; \|\widetilde{\mathbf{A}}\boldsymbol{\beta}\|_\infty \le \lambda, \quad \langle \mathbf{b}, \boldsymbol{\beta}\rangle = 0, \tag{A49}$$

where $\widetilde{\mathbf{A}} = [b_1\mathbf{a}_1, \ldots, b_n\mathbf{a}_n]$. It is easy to see there exists a $\lambda_{\max}$ such that (A48) and (A49) have an analytical solution for any $\lambda \ge \lambda_{\max}$. Specifically, for any $\lambda \ge \lambda_{\max}$, we have

$$\mathbf{w}_\lambda^\star = \mathbf{0}, \quad c_\lambda^\star = 0, \quad (\beta_\lambda)_i^\star = \begin{cases} \frac{n^-}{n} & \text{if } \; b_i = 1, \\ \frac{n^+}{n} & \text{if } \; b_i = -1 \end{cases} \quad \forall i = 1,\ldots,n.$$

The value of $\lambda_{\max}$ can be computed by $\lambda_{\max} = \|\widetilde{\mathbf{A}}\boldsymbol{\beta}^\star_{\lambda_{\max}}\|_\infty$. In our experiments, we set the value of λ as $\lambda = 0.05\lambda_{\max}$ for both datasets. Consequently, we obtain $\lambda = 5.751$ and $\lambda = 6.892$ for *MDS* and *RCV1*, respectively. For initialization, we set $\mathbf{w}_0^j = \mathbf{0}, v_0^j = 0, \mathbf{u}_0^j = \mathbf{0}, \forall j = 1,\ldots,J$ and $\mathbf{z}_0 = \mathbf{0}$.