Supplemental Material to "An Optimal Algorithm for Stochastic Three-Composite Optimization"

Renbo Zhao, William B. Haskell, Vincent Y. F. Tan

S-1 Lemmas

We provide some lemmas that will be used in proving Proposition 1.

Lemma S-1. *Let $\{\vartheta_k\}_{k\in\mathbb{Z}^+}$ and $\{\vartheta'_k\}_{k\in\mathbb{Z}^+}$ be any positive sequences and $\mathcal{X}\subseteq\mathbb{R}^d$ be convex and closed. Define a sequence of iterates $\{\mathbf{x}_\natural^k\}_{k\in\mathbb{Z}^+}$ such that $\mathbf{x}_\natural^0 \triangleq \mathbf{x}^0$ and*

$$\mathbf{x}_\natural^{k+1} \triangleq \Pi_{\mathcal{X}}\left[\mathbf{x}_\natural^k + \vartheta_k \boldsymbol{\varepsilon}^k\right],\ \forall\, k\in\mathbb{Z}^+,$$

where $\Pi_{\mathcal{X}}$ denotes the Euclidean projection onto $\mathcal{X}$ and the sequence $\{\boldsymbol{\varepsilon}^k\}_{k\in\mathbb{Z}^+}$ is defined in Assumption 1. Then

$$\sum_{k=0}^{K-1}\vartheta'_k\langle\boldsymbol{\varepsilon}^k,\mathbf{x}-\mathbf{x}_\natural^k\rangle \le \frac{\vartheta'_0}{2\vartheta_0}\|\mathbf{x}_\natural^0-\mathbf{x}\|^2 + \sum_{k=1}^{K-1}\left(\frac{\vartheta'_k}{2\vartheta_k}-\frac{\vartheta'_{k-1}}{2\vartheta_{k-1}}\right)\|\mathbf{x}_\natural^k-\mathbf{x}\|^2 + \sum_{k=0}^{K-1}\frac{\vartheta'_k\vartheta_k}{2}\|\boldsymbol{\varepsilon}_k\|^2,\ \forall\, K\in\mathbb{N}. \quad \text{(S-1)}$$

Proof. By the nonexpansiveness of $\Pi_{\mathcal{X}}$, for any $\mathbf{x}\in\mathcal{X}$ and $k\in\mathbb{Z}^+$, we have

$$\begin{aligned}\frac{1}{2}\|\mathbf{x}_\natural^{k+1}-\mathbf{x}\|^2 &= \frac{1}{2}\|\Pi_{\mathcal{X}}[\mathbf{x}_\natural^k+\vartheta_k\boldsymbol{\varepsilon}^k]-\Pi_{\mathcal{X}}[\mathbf{x}]\|^2\\ &\le \frac{1}{2}\|\mathbf{x}_\natural^k-\mathbf{x}+\vartheta_k\boldsymbol{\varepsilon}^k\|^2 = \frac{1}{2}\|\mathbf{x}_\natural^k-\mathbf{x}\|^2+\frac{\vartheta_k^2}{2}\|\boldsymbol{\varepsilon}^k\|^2+\vartheta_k\langle\mathbf{x}_\natural^k-\mathbf{x},\boldsymbol{\varepsilon}^k\rangle. \end{aligned}\quad \text{(S-2)}$$

Now, multiply both sides of (S-2) by ϑ'_k/ϑ_k and telescope over $k=0,\ldots,K-1$, we have

$$\begin{aligned}0 &\le \frac{\vartheta'_{K-1}}{2\vartheta_{K-1}}\|\mathbf{x}^K-\mathbf{x}\|^2\\ &\le \frac{\vartheta'_0}{2\vartheta_0}\|\mathbf{x}_\natural^0-\mathbf{x}\|^2+\sum_{k=1}^{K-1}\left(\frac{\vartheta'_k}{2\vartheta_k}-\frac{\vartheta'_{k-1}}{2\vartheta_{k-1}}\right)\|\mathbf{x}_\natural^k-\mathbf{x}\|^2+\sum_{k=0}^{K-1}\frac{\vartheta'_k\vartheta_k}{2}\|\boldsymbol{\varepsilon}^k\|^2+\sum_{k=0}^{K-1}\vartheta'_k\langle\boldsymbol{\varepsilon}^k,\mathbf{x}_\natural^k-\mathbf{x}\rangle.\end{aligned}$$

After rearranging, we arrive at (S-1). □

Lemma S-2. *Choose the input sequences $\{\beta_k\}_{k\in\mathbb{Z}^+}$, $\{\alpha_k\}_{k\in\mathbb{Z}^+}$, $\{\tau_k\}_{k\in\mathbb{Z}^+}$ and $\{\theta_k\}_{k\in\mathbb{Z}^+}$ in Algorithm 1 as in Section 2. If* $\mathbf{dom}\, g$ *and* $\mathbf{dom}\, h^*$ *are closed and bounded, then for any $K\in\mathbb{N}$,*

$$\begin{aligned}G(\overline{\mathbf{x}}^K,\overline{\mathbf{y}}^K) \le& \frac{1}{\beta_{K-1}\tau_{K-1}}D_g^2+\frac{1}{2\beta_{K-1}\alpha_{K-1}}D_{h^*}^2\\ &+\frac{1}{\beta_{K-1}\gamma_{K-1}}\sum_{k=0}^{K-1}\gamma_k\langle\boldsymbol{\varepsilon}^k,\mathbf{x}_\natural^k-\mathbf{x}^k\rangle+\frac{1+\zeta}{2\zeta\beta_{K-1}\gamma_{K-1}}\sum_{k=0}^{K-1}\gamma_k\tau_k\|\boldsymbol{\varepsilon}^k\|^2 \text{ a.s.}\end{aligned}\quad \text{(S-3)}$$

Proof. See Section S-4. □

Lemma S-3. *Let $\{\overline{\sigma}_k\}_{k\in\mathbb{Z}^+}$ be any positive sequence. Let $\{\delta_k\}_{k\in\mathbb{Z}^+}\subseteq\mathbb{R}$ be an martingale difference sequence (MDS) adapted to a filtration $\{\mathcal{F}'_k\}_{k\in\mathbb{Z}^+}$ such that for any $k\in\mathbb{Z}^+$, $\mathbb{E}[\delta_k\,|\,\mathcal{F}'_k]=0$ and $\mathbb{E}[\exp\{\delta_k^2/\overline{\sigma}_k^2\}\,|\,\mathcal{F}'_k]\le\exp\{1\}$ a.s. Then for any $p>0$ and $K\in\mathbb{N}$,*

$$\Pr\left(\sum_{k=0}^{K-1}\delta_i > p\sqrt{\sum_{k=0}^{K-1}\overline{\sigma}_i^2}\right)\le\exp\left\{-\frac{p^2}{4}\right\}. \quad \text{(S-4)}$$

Proof. See [1, Section 6]. □

S-2 Proof of Proposition 1

By the independence of ε^k and $\{\mathbf{x}_\natural^k, \mathbf{x}^k\}$ for any $k \in \mathbb{Z}^+$ and (A1), we have

$$\mathbb{E}_{\boldsymbol{\xi}^k}\left[\langle \varepsilon^k, \mathbf{x}_\natural^k - \mathbf{x}^k\rangle | \mathcal{F}_k\right] = 0, \ \forall\, K \in \mathbb{N}. \tag{S-5}$$

Therefore,

$$\mathbb{E}_{\Xi_K}\left[\sum_{k=0}^{K-1} \gamma_k \langle \varepsilon^k, \mathbf{x}_\natural^k - \mathbf{x}^k\rangle\right] = \sum_{k=0}^{K-1} \gamma_k \mathbb{E}\left[\mathbb{E}_{\boldsymbol{\xi}^k}\left[\langle \varepsilon^k, \mathbf{x}_\natural^k - \mathbf{x}^k\rangle | \mathcal{F}_k\right]\right] = 0. \tag{S-6}$$

By (A2), we also have

$$\mathbb{E}_{\Xi_K}\left[\sum_{k=0}^{K-1} \gamma_k \tau_k \|\varepsilon^k\|^2\right] = \sum_{k=0}^{K-1} \gamma_k \tau_k \mathbb{E}\left[\mathbb{E}_{\boldsymbol{\xi}^k}\left[\|\varepsilon^k\|^2 | \mathcal{F}_k\right]\right] \le \left(\sum_{k=0}^{K-1} \gamma_k \tau_k\right) \sigma^2. \tag{S-7}$$

Therefore, by combining (S-3), (S-6) and (S-7), we obtain (22).

We next prove (23). By (S-5), (A2) and the boundedness of $\mathbf{dom}\, g$, we see that $\{\gamma_k \langle \varepsilon^k, \mathbf{x}_\natural^k - \mathbf{x}^k\rangle\}_{k\in\mathbb{Z}^+}$ is a MDS adapted to $\{\mathcal{F}_k\}_{k\in\mathbb{Z}^+}$. In addition, by Cauchy-Schwartz and (A3),

$$\mathbb{E}_{\boldsymbol{\xi}^k}\left[\exp\left\{\frac{\gamma_k^2 |\langle \varepsilon^k, \mathbf{x}_\natural^k - \mathbf{x}^k\rangle|^2}{4\gamma_k^2 \sigma^2 D_g^2}\right\} \middle| \mathcal{F}_k\right] \le \exp\{1\}. \tag{S-8}$$

Then we invoke Lemma S-3 to obtain that for any $p > 0$,

$$\Pr\left\{\frac{1}{\beta_{K-1}\gamma_{K-1}} \sum_{k=0}^{K-1} \gamma_k \langle \varepsilon^k, \mathbf{x}_\natural^k - \mathbf{x}^k\rangle > \frac{2p\sigma D_g}{\beta_{K-1}\gamma_{K-1}} \sqrt{\sum_{k=0}^{K-1} \gamma_k^2}\right\} \tag{S-9}$$

$$= \Pr\left\{\sum_{k=0}^{K-1} \gamma_k \langle \varepsilon^k, \mathbf{x}_\natural^k - \mathbf{x}^k\rangle > p\sqrt{\sum_{k=0}^{K-1} 4\gamma_k^2 \sigma^2 D_g^2}\right\} \le \exp\{-p^2/4\}. \tag{S-10}$$

Recall from Proposition 1 that $\Gamma_K = \sum_{k=0}^{K-1} \gamma_k \tau_k$, for any $K \in \mathbb{N}$. Then by Jensen's inequality and (A3), for any $p' > 0$,

$$\mathbb{E}_{\Xi_K}\left[\exp\left\{\frac{p'}{2\Gamma_K} \sum_{k=0}^{K-1} \gamma_k \tau_k \frac{\|\varepsilon^k\|^2}{\sigma^2}\right\}\right] \le \frac{1}{\Gamma_K} \sum_{k=0}^{K-1} \gamma_k \tau_k \mathbb{E}\left[\mathbb{E}_{\boldsymbol{\xi}_k}\left[\exp\left\{\frac{p'\|\varepsilon^k\|^2}{2\sigma^2}\right\} \middle| \mathcal{F}_k\right]\right] \le \exp\{p'/2 + (p')^2/4\}. \tag{S-11}$$

Therefore, for any $p' > 0$,

$$\begin{aligned}
&\Pr\left\{\frac{1+\zeta}{2\zeta\beta_{K-1}\gamma_{K-1}} \sum_{k=0}^{K-1} \gamma_k \tau_k \|\varepsilon^k\|^2 > (1+p')\frac{\sigma^2(1+\zeta)}{2\zeta\beta_{K-1}\gamma_{K-1}} \Gamma_K\right\} \\
&= \Pr\left\{\exp\left\{\frac{p'}{2\Gamma_K} \sum_{k=0}^{K-1} \gamma_k \tau_k \frac{\|\varepsilon^k\|^2}{\sigma^2}\right\} > \exp\{p'(1+p')/2\}\right\} \\
&\le \mathbb{E}_{\Xi_K}\left[\exp\left\{\frac{p'}{2\Gamma_K} \sum_{k=0}^{K-1} \gamma_k \tau_k \frac{\|\varepsilon^k\|^2}{\sigma^2}\right\}\right] \exp\{-p'(1+p')/2\} \le \exp\{-(p')^2/4\},
\end{aligned} \tag{S-12}$$

where (S-12) follows from Markov's inequality and (S-11).

Recall from Proposition 1 that $\Gamma'_K \triangleq (\sum_{k=0}^{K-1} \gamma_k^2)^{1/2}$. Based on (S-3), (S-10) and (S-12), we have that for any $p, p' > 0$,

$$\begin{aligned}
\Pr\Bigg\{G(\overline{\mathbf{x}}^K, \overline{\mathbf{y}}^K) > \frac{D_g^2}{\beta_{K-1}\tau_{K-1}} + \frac{D_{h^*}^2}{2\beta_{K-1}\alpha_{K-1}} + \frac{2p\sigma D_g}{\beta_{K-1}\gamma_{K-1}} \Gamma'_K \\
+ (1+p')\frac{\sigma^2(1+\zeta)}{2\zeta\beta_{K-1}\gamma_{K-1}} \Gamma_K\Bigg\} \le \exp\{-p^2/4\} + \exp\{-(p')^2/4\}.
\end{aligned} \tag{S-13}$$

Taking $p = p' = 2\sqrt{\log(2/\delta)}$, we then complete the proof.

S-3 Proof of Theorem 1

First, we easily see that the conditions on $\{\beta_k\}_{k\in\mathbb{Z}^+}$, $\{\alpha_k\}_{k\in\mathbb{Z}^+}$, $\{\tau_k\}_{k\in\mathbb{Z}^+}$ and $\{\theta_k\}_{k\in\mathbb{Z}^+}$ in Proposition 1 are all satisfied with $\zeta = 1/2$. In particular, for any $k \in \mathbb{N}$,

$$\frac{1}{2\tau_{k-1}} - \frac{L}{\beta_{k-1}} - B^2\alpha_{k-1} = \frac{2}{k+1}L + \frac{\rho\sigma}{2}\sqrt{k+1} - \frac{2(k+1)}{k(k+3)}L = \frac{2(k-1)}{k(k+1)(k+3)}L + \frac{\rho\sigma}{2}\sqrt{k+1} \geq 0.$$

From (22), we have

$$\begin{aligned}\mathbb{E}_{\Xi_K}\left[G(\overline{\mathbf{x}}^K, \overline{\mathbf{y}}^K)\right] &\overset{\text{(a)}}{\leq} \frac{2(K+1)}{K(K+3)}\left\{\left(\frac{4L}{K+1} + 2\rho' B + \rho\sigma\sqrt{K+1}\right)D_g^2 + \frac{B}{2\rho'}D_{h^*}^2 + \frac{3}{2\rho K}\left(\sum_{k=1}^{K}\sqrt{k}\right)\sigma\right\}\\ &\overset{\text{(b)}}{\leq} \frac{8L}{K(K+3)}D_g^2 + \frac{4(K+1)}{K(K+3)}\left(\rho' D_g^2 + \frac{D_{h^*}^2}{4\rho'}\right)B + \frac{2(K+1)^{3/2}}{K(K+3)}\left(\rho D_g^2 + \frac{2}{\rho}\right)\sigma.\end{aligned}$$

where in (a) we use $\tau_k \leq 1/(\rho\sigma\sqrt{k+1})$ for any $k \in \mathbb{Z}^+$ and in (b) we use $\sum_{k=1}^{K}\sqrt{k} \leq (2/3)(K+1)^{3/2}$ for any $K \in \mathbb{N}$.

In addition, in (23), we have

$$\begin{aligned}&\frac{4\sqrt{\log(2/\delta)}D_g}{\beta_{K-1}\gamma_{K-1}}\sqrt{\sum_{k=0}^{K-1}\gamma_k^2\sigma + \frac{(1+2\sqrt{\log(2/\delta)})(1+\zeta)}{2\zeta\beta_{K-1}\gamma_{K-1}}\left(\sum_{k=0}^{K-1}\gamma_k\tau_k\right)\sigma^2}\\ &\overset{\text{(c)}}{\leq} \frac{2(K+1)}{K(K+3)}\left\{4\sqrt{K\log(2/\delta)}D_g + \frac{1+2\sqrt{\log(2/\delta)}}{\rho K}(K+1)^{3/2}\right\}\sigma\\ &\leq \frac{8(K+1)^{3/2}}{K(K+3)}\left(\sqrt{\log(2/\delta)}D_g + \frac{1/2+\sqrt{\log(2/\delta)}}{\rho}\right)\sigma \overset{\text{(d)}}{\leq} \frac{16}{\sqrt{K+3}}\left(D_g + \frac{2}{\rho}\right)\sqrt{\log(2/\delta)}\sigma,\end{aligned} \tag{S-14}$$

where in (c) we use $\sum_{k=1}^{K} k^2 \leq K^3$ for any $K \in \mathbb{N}$ and in (d) we use $\delta \in (0,1)$.

S-4 Proof of Lemma S-2

For any $k \in \mathbb{N}$, from steps (5) and (6) and the first-order optimality conditions, we have

$$\alpha_k^{-1}(\mathbf{y}^k - \mathbf{y}^{k+1}) + \mathbf{A}\mathbf{z}^k \in \partial h^*(\mathbf{y}^{k+1}), \tag{S-15}$$

$$\tau_k^{-1}(\mathbf{x}^k - \mathbf{x}^{k+1}) - (\mathbf{v}^k + \mathbf{A}^T\mathbf{y}^{k+1}) \in \partial g(\mathbf{x}^{k+1}). \tag{S-16}$$

Using the definitions of subdifferential and law of cosines, for any $\mathbf{x} \in \mathbb{R}^d$ and $\mathbf{y} \in \mathbb{R}^m$,

$$h^*(\mathbf{y}) \geq h^*(\mathbf{y}^{k+1}) + \frac{1}{2\alpha_k}\left(\|\mathbf{y}^k - \mathbf{y}^{k+1}\|^2 + \|\mathbf{y}^{k+1} - \mathbf{y}\|^2 - \|\mathbf{y}^k - \mathbf{y}\|^2\right) + \langle \mathbf{A}\mathbf{z}^k, \mathbf{y} - \mathbf{y}^{k+1}\rangle, \tag{S-17}$$

$$\begin{aligned}g(\mathbf{x}) \geq\ & g(\mathbf{x}^{k+1}) + \frac{1}{2\tau_k}\left(\|\mathbf{x}^k - \mathbf{x}^{k+1}\|^2 + \|\mathbf{x}^{k+1} - \mathbf{x}\|^2 - \|\mathbf{x}^k - \mathbf{x}\|^2\right)\\ &+ \langle \boldsymbol{\varepsilon}^k, \mathbf{x}^{k+1} - \mathbf{x}\rangle + \langle \nabla f(\widetilde{\mathbf{x}}^k), \mathbf{x}^{k+1} - \mathbf{x}\rangle + \langle \mathbf{A}(\mathbf{x}^{k+1} - \mathbf{x}), \mathbf{y}^{k+1}\rangle.\end{aligned} \tag{S-18}$$

In addition, from steps (4) and (8), we have $\overline{\mathbf{x}}^{k+1} - \widetilde{\mathbf{x}}^k = \beta_k^{-1}(\mathbf{x}^{k+1} - \mathbf{x}^k)$ and for any $\mathbf{x} \in \mathbb{R}^d$,

$$\begin{aligned}
&(\beta_k - 1) f(\overline{\mathbf{x}}^k) + f(\mathbf{x}) \\
&\geq (\beta_k - 1)\left(f(\widetilde{\mathbf{x}}^k) + \langle \nabla f(\widetilde{\mathbf{x}}^k), \overline{\mathbf{x}}^k - \widetilde{\mathbf{x}}^k\rangle\right) + f(\widetilde{\mathbf{x}}^k) + \langle \nabla f(\widetilde{\mathbf{x}}^k), \mathbf{x} - \widetilde{\mathbf{x}}^k\rangle \\
&= \beta_k f(\widetilde{\mathbf{x}}^k) + (\beta_k - 1)\langle \nabla f(\widetilde{\mathbf{x}}^k), \overline{\mathbf{x}}^k - \widetilde{\mathbf{x}}^k\rangle + \langle \nabla f(\widetilde{\mathbf{x}}^k), \mathbf{x}^{k+1} - \widetilde{\mathbf{x}}^k\rangle + \langle \nabla f(\widetilde{\mathbf{x}}^k), \mathbf{x} - \mathbf{x}^{k+1}\rangle \\
&= \beta_k \left(f(\widetilde{\mathbf{x}}^k) + (1 - \beta_k^{-1})\langle \nabla f(\widetilde{\mathbf{x}}^k), \overline{\mathbf{x}}^k - \widetilde{\mathbf{x}}^k\rangle + \beta_k^{-1}\langle \nabla f(\widetilde{\mathbf{x}}^k), \mathbf{x}^{k+1} - \widetilde{\mathbf{x}}^k\rangle\right) + \langle \nabla f(\widetilde{\mathbf{x}}^k), \mathbf{x} - \mathbf{x}^{k+1}\rangle \\
&\geq \beta_k \left(f(\overline{\mathbf{x}}^{k+1}) - (L/2)\|\overline{\mathbf{x}}^{k+1} - \widetilde{\mathbf{x}}^k\|^2\right) + \langle \nabla f(\widetilde{\mathbf{x}}^k), \mathbf{x} - \mathbf{x}^{k+1}\rangle \\
&= \beta_k f(\overline{\mathbf{x}}^{k+1}) - \frac{L}{2\beta_k}\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 + \langle \nabla f(\widetilde{\mathbf{x}}^k), \mathbf{x} - \mathbf{x}^{k+1}\rangle. \qquad \text{(S-19)}
\end{aligned}$$

From (S-19), we immediately see that

$$\beta_k(f(\overline{\mathbf{x}}^{k+1}) - f(\mathbf{x})) \leq (\beta_k - 1)(f(\overline{\mathbf{x}}^k) - f(\mathbf{x})) + \frac{L}{2\beta_k}\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 + \langle \nabla f(\widetilde{\mathbf{x}}^k), \mathbf{x}^{k+1} - \mathbf{x}\rangle. \qquad \text{(S-20)}$$

Also, by Jensen's inequality, we have for any $\mathbf{x} \in \mathbb{R}^d$ and $\mathbf{y} \in \mathbb{R}^m$,

$$g(\overline{\mathbf{x}}^{k+1}) - g(\mathbf{x}) \leq \beta_k^{-1}(g(\mathbf{x}^{k+1}) - g(\mathbf{x})) + (1 - \beta_k^{-1})(g(\overline{\mathbf{x}}^k) - g(\mathbf{x})), \qquad \text{(S-21)}$$
$$h^*(\overline{\mathbf{y}}^{k+1}) - h^*(\mathbf{y}) \leq \beta_k^{-1}(h^*(\mathbf{y}^{k+1}) - h^*(\mathbf{y})) + (1 - \beta_k^{-1})(h^*(\overline{\mathbf{y}}^k) - h^*(\mathbf{y})). \qquad \text{(S-22)}$$

For convenience, for any $k \in \mathbb{Z}^+$, $\mathbf{x} \in \mathbb{R}^d$ and $\mathbf{y} \in \mathbb{R}^m$, define $\overline{\mathbf{w}}^k \triangleq (\overline{\mathbf{x}}^k, \overline{\mathbf{y}}^k)$ and $\mathbf{w} \triangleq (\mathbf{x}, \mathbf{y})$. Accordingly, define

$$\begin{aligned}
Q(\overline{\mathbf{w}}^k, \mathbf{w}) &\triangleq S(\overline{\mathbf{x}}^k, \mathbf{y}) - S(\mathbf{x}, \overline{\mathbf{y}}^k) \\
&= (f(\overline{\mathbf{x}}^k) - f(\mathbf{x})) + (g(\overline{\mathbf{x}}^k) - g(\mathbf{x})) + (\langle \mathbf{A}\overline{\mathbf{x}}^k, \mathbf{y}\rangle - \langle \mathbf{A}\mathbf{x}, \overline{\mathbf{y}}^k\rangle) + (h^*(\overline{\mathbf{y}}^k) - h^*(\mathbf{y})). \qquad \text{(S-23)}
\end{aligned}$$

Then for any $\mathbf{x} \in \mathbb{R}^d$ and $\mathbf{y} \in \mathbb{R}^m$,

$$\begin{aligned}
&\beta_k Q(\overline{\mathbf{w}}^{k+1}, \mathbf{w}) \\
&\overset{\text{(S-20),(8),(9)}}{\leq} (\beta_k - 1)(f(\overline{\mathbf{x}}^k) - f(\mathbf{x})) + \frac{L}{2\beta_k}\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 + \langle \nabla f(\widetilde{\mathbf{x}}^k), \mathbf{x}^{k+1} - \mathbf{x}\rangle \\
&\quad + (\beta_k - 1)(\langle \mathbf{A}\overline{\mathbf{x}}^k, \mathbf{y}\rangle - \langle \mathbf{A}\mathbf{x}, \overline{\mathbf{y}}^k\rangle) + (\langle \mathbf{A}\mathbf{x}^{k+1}, \mathbf{y}\rangle - \langle \mathbf{A}\mathbf{x}, \mathbf{y}^{k+1}\rangle) \\
&\quad + \beta_k(g(\overline{\mathbf{x}}^{k+1}) - g(\mathbf{x})) + \beta_k(h^*(\overline{\mathbf{y}}^{k+1}) - h^*(\mathbf{y})) \\
&\overset{\text{(S-21),(S-22)}}{\leq} (\beta_k - 1)(f(\overline{\mathbf{x}}^k) - f(\mathbf{x})) + (\beta_k - 1)(g(\overline{\mathbf{x}}^k) - g(\mathbf{x})) + (\beta_k - 1)(\langle \mathbf{A}\overline{\mathbf{x}}^k, \mathbf{y}\rangle - \langle \mathbf{A}\mathbf{x}, \overline{\mathbf{y}}^k\rangle) \\
&\quad + (\beta_k - 1)(h^*(\overline{\mathbf{y}}^k) - h^*(\mathbf{y})) + \frac{L}{2\beta_k}\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 + \langle \nabla f(\widetilde{\mathbf{x}}^k), \mathbf{x}^{k+1} - \mathbf{x}\rangle \\
&\quad + (\langle \mathbf{A}\mathbf{x}^{k+1}, \mathbf{y}\rangle - \langle \mathbf{A}\mathbf{x}, \mathbf{y}^{k+1}\rangle) + (g(\mathbf{x}^{k+1}) - g(\mathbf{x})) + (h^*(\mathbf{y}^{k+1}) - h^*(\mathbf{y})) \\
&\overset{\text{(S-23),(S-17),(S-18)}}{\leq} (\beta_k - 1)Q(\overline{\mathbf{w}}^k, \mathbf{w}) + \frac{L}{2\beta_k}\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 + \langle \mathbf{A}\mathbf{x}^{k+1}, \mathbf{y} - \mathbf{y}^{k+1}\rangle \\
&\quad + \frac{1}{2\tau_k}\left(\|\mathbf{x}^k - \mathbf{x}\|^2 - \|\mathbf{x}^k - \mathbf{x}^{k+1}\|^2 - \|\mathbf{x}^{k+1} - \mathbf{x}\|^2\right) + \langle \boldsymbol{\varepsilon}^k, \mathbf{x} - \mathbf{x}^{k+1}\rangle \\
&\quad + \frac{1}{2\alpha_k}\left(\|\mathbf{y}^k - \mathbf{y}\|^2 - \|\mathbf{y}^k - \mathbf{y}^{k+1}\|^2 - \|\mathbf{y}^{k+1} - \mathbf{y}\|^2\right) + \langle \mathbf{A}\mathbf{z}^k, \mathbf{y}^{k+1} - \mathbf{y}\rangle \\
&\overset{(7)}{\leq} (\beta_k - 1)Q(\overline{\mathbf{w}}^k, \mathbf{w}) + \left(\frac{L}{2\beta_k} - \frac{1}{2\tau_k}\right)\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 + \frac{1}{2\tau_k}\left(\|\mathbf{x}^k - \mathbf{x}\|^2 - \|\mathbf{x}^{k+1} - \mathbf{x}\|^2\right) \\
&\quad + \langle \mathbf{A}(\mathbf{x}^k - \mathbf{x}^{k+1}), \mathbf{y}^{k+1} - \mathbf{y}\rangle + \theta_k\langle \mathbf{A}(\mathbf{x}^k - \mathbf{x}^{k-1}), \mathbf{y}^k - \mathbf{y}\rangle + \theta_k\langle \mathbf{A}(\mathbf{x}^k - \mathbf{x}^{k-1}), \mathbf{y}^{k+1} - \mathbf{y}^k\rangle \\
&\quad + \frac{1}{2\alpha_k}\left(\|\mathbf{y}^k - \mathbf{y}\|^2 - \|\mathbf{y}^{k+1} - \mathbf{y}\|^2\right) - \frac{1}{2\alpha_k}\|\mathbf{y}^k - \mathbf{y}^{k+1}\|^2 + \langle \boldsymbol{\varepsilon}^k, \mathbf{x} - \mathbf{x}^{k+1}\rangle. \qquad \text{(S-24)}
\end{aligned}$$

Define $\gamma_{-1} \triangleq 1$. From the definition of $\{\gamma_k\}_{k\in\mathbb{Z}^+}$, we see that $\gamma_{k-1} = \gamma_k\theta_k$, for any $k \in \mathbb{Z}^+$. Now, we multiply both sides of (S-24) by γ_k and use condition (19) to obtain

$$\begin{aligned}
&\beta_k\gamma_k Q(\overline{\mathbf{w}}^{k+1}, \mathbf{w}) \\
&\leq \beta_{k-1}\gamma_{k-1}Q(\overline{\mathbf{w}}^k, \mathbf{w}) + \gamma_k\left(\frac{L}{2\beta_k} - \frac{1}{2\tau_k}\right)\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 + \frac{\gamma_k}{2\tau_k}\left(\|\mathbf{x}^k - \mathbf{x}\|^2 - \|\mathbf{x}^{k+1} - \mathbf{x}\|^2\right) \\
&+ \gamma_k\langle \mathbf{A}(\mathbf{x}^k - \mathbf{x}^{k+1}), \mathbf{y}^{k+1} - \mathbf{y}\rangle + \gamma_{k-1}\langle \mathbf{A}(\mathbf{x}^k - \mathbf{x}^{k-1}), \mathbf{y}^k - \mathbf{y}\rangle + \gamma_{k-1}\langle \mathbf{A}(\mathbf{x}^k - \mathbf{x}^{k-1}), \mathbf{y}^{k+1} - \mathbf{y}^k\rangle \\
&+ \frac{\gamma_k}{2\alpha_k}\left(\|\mathbf{y}^k - \mathbf{y}\|^2 - \|\mathbf{y}^{k+1} - \mathbf{y}\|^2\right) - \frac{\gamma_k}{2\alpha_k}\|\mathbf{y}^k - \mathbf{y}^{k+1}\|^2 + \gamma_k\langle \boldsymbol{\varepsilon}^k, \mathbf{x} - \mathbf{x}^{k+1}\rangle. \qquad \text{(S-25)}
\end{aligned}$$

By Young's inequality and that $|||\mathbf{A}||| = B$, we have

$$\gamma_{k-1}\langle \mathbf{A}(\mathbf{x}^k - \mathbf{x}^{k-1}), \mathbf{y}^{k+1} - \mathbf{y}^k\rangle \leq \frac{B^2\theta_k^2\alpha_k\gamma_k}{2}\|\mathbf{x}^k - \mathbf{x}^{k-1}\|^2 + \frac{\gamma_k}{2\alpha_k}\|\mathbf{y}^{k+1} - \mathbf{y}^k\|^2. \qquad \text{(S-26)}$$

Substituting (S-26) into (S-25), we have

$$\begin{aligned}
&\beta_k\gamma_k Q(\overline{\mathbf{w}}^{k+1}, \mathbf{w}) - \beta_{k-1}\gamma_{k-1}Q(\overline{\mathbf{w}}^k, \mathbf{w}) \\
&\leq \frac{B^2\theta_k^2\alpha_k\gamma_k}{2}\|\mathbf{x}^k - \mathbf{x}^{k-1}\|^2 - \gamma_k\left(\frac{1}{2\tau_k} - \frac{L}{2\beta_k}\right)\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 + \frac{\gamma_k}{2\tau_k}\left(\|\mathbf{x}^k - \mathbf{x}\|^2 - \|\mathbf{x}^{k+1} - \mathbf{x}\|^2\right) \\
&+ \gamma_{k-1}\langle \mathbf{A}(\mathbf{x}^k - \mathbf{x}^{k-1}), \mathbf{y}^k - \mathbf{y}\rangle - \gamma_k\langle \mathbf{A}(\mathbf{x}^{k+1} - \mathbf{x}^k), \mathbf{y}^{k+1} - \mathbf{y}\rangle \\
&+ \frac{\gamma_k}{2\alpha_k}\left(\|\mathbf{y}^k - \mathbf{y}\|^2 - \|\mathbf{y}^{k+1} - \mathbf{y}\|^2\right) + \gamma_k\langle \boldsymbol{\varepsilon}^k, \mathbf{x} - \mathbf{x}^{k+1}\rangle. \qquad \text{(S-27)}
\end{aligned}$$

For any fixed $K \in \mathbb{N}$, we then telescope (S-27) over $k = 0, \ldots, K-1$ to obtain

$$\begin{aligned}
&\beta_{K-1}\gamma_{K-1}Q(\overline{\mathbf{w}}^K, \mathbf{w}) - \beta_{-1}\gamma_{-1}Q(\overline{\mathbf{w}}^0, \mathbf{w}) \leq \frac{B^2\theta_0^2\alpha_0\gamma_0}{2}\|\mathbf{x}^0 - \mathbf{x}^{-1}\|^2 \\
&+ \sum_{k=1}^{K-1}\frac{\gamma_{k-1}}{2}\left(B^2\theta_k\alpha_k + \frac{L}{\beta_{k-1}} - \frac{1}{\tau_{k-1}}\right)\|\mathbf{x}^k - \mathbf{x}^{k-1}\|^2 - \frac{\gamma_{K-1}}{2}\left(\frac{1}{\tau_{K-1}} - \frac{L}{\beta_{K-1}}\right)\|\mathbf{x}^K - \mathbf{x}^{K-1}\|^2 \\
&+ \frac{\gamma_0}{2\tau_0}\|\mathbf{x}^0 - \mathbf{x}\|^2 + \sum_{k=1}^{K-1}\left(\frac{\gamma_k}{2\tau_k} - \frac{\gamma_{k-1}}{2\tau_{k-1}}\right)\|\mathbf{x}^k - \mathbf{x}\|^2 - \frac{\gamma_{K-1}}{2\tau_{K-1}}\|\mathbf{x}^K - \mathbf{x}\|^2 \\
&+ \frac{\gamma_0}{2\alpha_0}\|\mathbf{y}^0 - \mathbf{y}\|^2 + \sum_{k=1}^{K-1}\left(\frac{\gamma_k}{2\alpha_k} - \frac{\gamma_{k-1}}{2\alpha_{k-1}}\right)\|\mathbf{y}^k - \mathbf{y}\|^2 - \frac{\gamma_{K-1}}{2\alpha_{K-1}}\|\mathbf{y}^K - \mathbf{y}\|^2 \\
&+ \gamma_{-1}\langle \mathbf{A}(\mathbf{x}^0 - \mathbf{x}^{-1}), \mathbf{y}^0 - \mathbf{y}\rangle - \gamma_{K-1}\langle \mathbf{A}(\mathbf{x}^K - \mathbf{x}^{K-1}), \mathbf{y}^K - \mathbf{y}\rangle + \sum_{k=0}^{K-1}\gamma_k\langle \boldsymbol{\varepsilon}^k, \mathbf{x} - \mathbf{x}^{k+1}\rangle. \qquad \text{(S-28)}
\end{aligned}$$

Since $\mathbf{z}^0 = \mathbf{x}^0$ and $\theta_0 > 0$, we have $\mathbf{x}^{-1} = \mathbf{x}^0$. In addition, from (19), we see that $\beta_{-1}\gamma_{-1} = 0$. By Young's inequality,

$$-\gamma_{K-1}\langle \mathbf{A}(\mathbf{x}^K - \mathbf{x}^{K-1}), \mathbf{y}^K - \mathbf{y}\rangle \leq \frac{B^2\alpha_{K-1}\gamma_{K-1}}{2}\|\mathbf{x}^K - \mathbf{x}^{K-1}\|^2 + \frac{\gamma_{K-1}}{2\alpha_{K-1}}\|\mathbf{y}^K - \mathbf{y}\|^2. \qquad \text{(S-29)}$$

By condition (20) and the boundedness of $\mathbf{dom}\, g$ and $\mathbf{dom}\, h^*$, we have

$$\frac{\gamma_0}{2\tau_0}\|\mathbf{x}^0 - \mathbf{x}\|^2 + \sum_{k=1}^{K-1}\left(\frac{\gamma_k}{2\tau_k} - \frac{\gamma_{k-1}}{2\tau_{k-1}}\right)\|\mathbf{x}^k - \mathbf{x}\|^2 \leq \frac{\gamma_{K-1}}{2\tau_{K-1}}D_g^2, \qquad \text{(S-30)}$$

$$\frac{\gamma_0}{2\alpha_0}\|\mathbf{y}^0 - \mathbf{y}\|^2 + \sum_{k=1}^{K-1}\left(\frac{\gamma_k}{2\alpha_k} - \frac{\gamma_{k-1}}{2\alpha_{k-1}}\right)\|\mathbf{y}^k - \mathbf{y}\|^2 \leq \frac{\gamma_{K-1}}{2\alpha_{K-1}}D_{h^*}^2. \qquad \text{(S-31)}$$

By condition (20), we also have $\theta_k\alpha_k \le \alpha_{k-1}$, for any $k \in \mathbb{N}$. Thus, (S-28) now becomes

$$\begin{aligned}\beta_{K-1}\gamma_{K-1}Q(\overline{\mathbf{w}}^K, \mathbf{w}) \le & \sum_{k=1}^{K} \frac{\gamma_{k-1}}{2}\left(B^2\alpha_{k-1} + \frac{L}{\beta_{k-1}} - \frac{1}{\tau_{k-1}}\right)\|\mathbf{x}^k - \mathbf{x}^{k-1}\|^2 \\ & + \frac{\gamma_{K-1}}{2\tau_{K-1}}D_g^2 - \frac{\gamma_{K-1}}{2\tau_{K-1}}\|\mathbf{x}^K - \mathbf{x}\|^2 + \frac{\gamma_{K-1}}{2\alpha_{K-1}}D_{h^*}^2 + \sum_{k=0}^{K-1}\gamma_k\langle \boldsymbol{\varepsilon}^k, \mathbf{x} - \mathbf{x}^{k+1}\rangle. \end{aligned} \tag{S-32}$$

Now, we decompose the last term in (S-32) into three parts, i.e.,

$$\sum_{k=0}^{K-1}\gamma_k\langle \boldsymbol{\varepsilon}^k, \mathbf{x} - \mathbf{x}^{k+1}\rangle = \sum_{k=0}^{K-1}\gamma_k\langle \boldsymbol{\varepsilon}^k, \mathbf{x} - \mathbf{x}_\natural^{k}\rangle + \sum_{k=0}^{K-1}\gamma_k\langle \boldsymbol{\varepsilon}^k, \mathbf{x}_\natural^{k} - \mathbf{x}^k\rangle + \sum_{k=0}^{K-1}\gamma_k\langle \boldsymbol{\varepsilon}^k, \mathbf{x}^k - \mathbf{x}^{k+1}\rangle, \tag{S-33}$$

where $\{\mathbf{x}_\natural^k\}_{k\in\mathbb{Z}^+}$ is defined as in Lemma S-1 with $\vartheta_k = \tau_k$, for any $k \in \mathbb{Z}^+$. By Lemma S-1, we have

$$\begin{aligned}\sum_{k=0}^{K-1}\gamma_k\langle \boldsymbol{\varepsilon}^k, \mathbf{x} - \mathbf{x}_\natural^{k}\rangle &\le \frac{\gamma_0}{2\tau_0}\|\mathbf{x}_\natural^0 - \mathbf{x}\|^2 + \sum_{k=1}^{K-1}\left(\frac{\gamma_k}{2\tau_k} - \frac{\gamma_{k-1}}{2\tau_{k-1}}\right)\|\mathbf{x}_\natural^k - \mathbf{x}\|^2 + \sum_{k=0}^{K-1}\frac{\gamma_k\tau_k}{2}\|\boldsymbol{\varepsilon}^k\|^2 \\ &\le \frac{\gamma_{K-1}}{2\tau_{K-1}}D_g^2 + \sum_{k=0}^{K-1}\frac{\gamma_k\tau_k}{2}\|\boldsymbol{\varepsilon}^k\|^2. \end{aligned} \tag{S-34}$$

By Young's inequality, for any $\zeta \in (0, 1)$,

$$\sum_{k=0}^{K-1}\gamma_k\langle \boldsymbol{\varepsilon}^k, \mathbf{x}^k - \mathbf{x}^{k+1}\rangle \le \sum_{k=0}^{K-1}\frac{\gamma_k\tau_k}{2\zeta}\|\boldsymbol{\varepsilon}^k\|^2 + \sum_{k=1}^{K}\frac{\zeta\gamma_{k-1}}{2\tau_{k-1}}\|\mathbf{x}^k - \mathbf{x}^{k-1}\|^2. \tag{S-35}$$

Substitute (S-33), (S-34) and (S-35) into (S-32), we then have

$$\begin{aligned}\beta_{K-1}\gamma_{K-1}Q(\overline{\mathbf{w}}^K, \mathbf{w}) \le & \sum_{k=1}^{K} \frac{\gamma_{k-1}}{2}\left(B^2\alpha_{k-1} + \frac{L}{\beta_{k-1}} - \frac{1-\zeta}{\tau_{k-1}}\right)\|\mathbf{x}^k - \mathbf{x}^{k-1}\|^2 \\ & + \frac{\gamma_{K-1}}{\tau_{K-1}}D_g^2 + \frac{\gamma_{K-1}}{2\alpha_{K-1}}D_{h^*}^2 + \sum_{k=0}^{K-1}\gamma_k\langle \boldsymbol{\varepsilon}^k, \mathbf{x}_\natural^k - \mathbf{x}^k\rangle + \frac{1+\zeta}{2\zeta}\sum_{k=0}^{K-1}\gamma_k\tau_k\|\boldsymbol{\varepsilon}^k\|^2. \end{aligned} \tag{S-36}$$

Apply condition (21) to (S-36) followed by taking supremum over $\mathbf{w} \in \mathbf{dom}\, g \times \mathbf{dom}\, h^*$, we then obtain (S-3).

S-5 Convergence Analysis of Algorithm 2

We now focus on Problem (16). For any $\mathbf{x}\in\mathbb{R}^d$ and $\widehat{\mathbf{y}}\in\mathbb{R}^{\mathbf{m}}$, define the primal-dual gap

$$\widehat{G}(\mathbf{x}, \mathbf{y}_1, \ldots, \mathbf{y}_p) \triangleq \sup_{(\mathbf{y}_1', \ldots, \mathbf{y}_p')\in \mathbf{dom}\, H^*} \widehat{S}(\mathbf{x}, \mathbf{y}_1', \ldots, \mathbf{y}_p') - \inf_{\mathbf{x}'\in\mathbf{dom}\, g} \widehat{S}(\mathbf{x}', \mathbf{y}_1, \ldots, \mathbf{y}_p).$$

Based on Theorem 1 (and Remark 9), we can obtain the following convergence results for Algorithm 2, by noting that $\mathbf{dom}\, H^* = \prod_{i=1}^p \mathbf{dom}\, h_i^*$, $D_{H^*} = (\sum_{i=1}^p D_{h_i^*}^2)^{1/2}$ and $\|\widehat{\mathbf{A}}\| = (\sum_{i=1}^p B_i^2)^{1/2} \triangleq \widehat{B}$.

Corollary S-1. *Let $\mathbf{dom}\, g$ be compact and $\mathbf{dom}\, h_i^*$ be bounded for each $i\in[p]$. In Algorithm 2, choose $\{\beta_k\}_{k\in\mathbb{Z}^+}$, $\{\alpha_k\}_{k\in\mathbb{Z}^+}$, $\{\tau_k\}_{k\in\mathbb{Z}^+}$ and $\{\theta_k\}_{k\in\mathbb{Z}^+}$ as in (11) and (12), and constants $\rho' = D_{H^*}/(2D_g)$ and $\rho = 1/D_g$. If (A1) and (A2) hold, then for any $K \in \mathbb{N}$,*

$$\mathbb{E}_{\Xi_K}\left[\widehat{G}(\overline{\mathbf{x}}^K, \overline{\mathbf{y}}_1^K, \ldots, \overline{\mathbf{y}}_p^K)\right] \le 8LD_g^2/[K(K+3)] + 4\widehat{B}D_gD_{H^*}/K + 12\sigma D_g/\sqrt{K+3}. \tag{S-37}$$

In addition, if (A1) and (A3) hold, then for any $\delta \in (0, 1)$,

$$\widehat{G}(\overline{\mathbf{x}}^K, \overline{\mathbf{y}}_1^K, \ldots, \overline{\mathbf{y}}_p^K) \le 8LD_g^2/[K(K+3)] + 4\widehat{B}D_gD_{H^*}/K + 32\sigma\sqrt{\log(2/\delta)}D_g/\sqrt{K+3}$$

w.p. at least $1-\delta$.

S-6 Proof of Important Steps in Section 4.2

We first prove step (41). From (30), we have

$$\begin{aligned}\boldsymbol{\omega}^{k+1} &= \underset{\boldsymbol{\omega}\in\mathbf{dom}\,h}{\arg\min}\; h(\boldsymbol{\omega}) - \langle \boldsymbol{\lambda}^k, \mathbf{A}\mathbf{u}^k - \boldsymbol{\omega}\rangle + \frac{\varrho}{2}\|\mathbf{A}\mathbf{u}^k - \boldsymbol{\omega}\|^2 \\ &= \underset{\boldsymbol{\omega}\in\mathbf{dom}\,h}{\arg\min}\; h(\boldsymbol{\omega}) + \frac{\varrho}{2}\|\mathbf{A}\mathbf{u}^k - \boldsymbol{\omega} - \boldsymbol{\lambda}^k/\varrho\|^2 \\ &= \mathbf{prox}_{h/\varrho}(\mathbf{A}\mathbf{u}^k - \boldsymbol{\lambda}^k/\varrho) \\ &= \mathbf{A}\mathbf{u}^k - \frac{1}{\varrho}(\boldsymbol{\lambda}^k + \mathbf{prox}_{\varrho h^*}(\varrho\mathbf{A}\mathbf{u}^k - \boldsymbol{\lambda}^k)) = \mathbf{A}\mathbf{u}^k - \frac{1}{\varrho}(\boldsymbol{\lambda}^k + \mathbf{y}_{\diamondsuit}^{k+1}),\end{aligned} \tag{S-38}$$

where in (S-38) we first use Moreau's identity in (17) and then the definition of $\mathbf{y}_{\diamondsuit}^{k+1}$ in (40).

Next, we show step (42). From (31), we have

$$\begin{aligned}\mathbf{u}^{k+1} &= \underset{\mathbf{u}\in\mathbf{dom}\,g}{\arg\min}\; g(\mathbf{u}) + \langle \mathbf{v}^k, \mathbf{u}-\mathbf{u}^k\rangle + \frac{r_k}{2\eta_k}\langle \mathbf{u}-\mathbf{u}^k, \mathbf{W}^k(\mathbf{u}-\mathbf{u}^k)\rangle + \frac{\varrho}{2}\|\mathbf{A}\mathbf{u} - \boldsymbol{\omega}^{k+1} - \boldsymbol{\lambda}^k/\varrho\|^2 \\ &\overset{\text{(a)}}{=} \underset{\mathbf{u}\in\mathbf{dom}\,g}{\arg\min}\; g(\mathbf{u}) + \langle \mathbf{v}^k, \mathbf{u}-\mathbf{u}^k\rangle + \frac{r_k}{2\eta_k}\langle \mathbf{u}-\mathbf{u}^k, \mathbf{W}^k(\mathbf{u}-\mathbf{u}^k)\rangle + \frac{\varrho}{2}\|\mathbf{A}(\mathbf{u}-\mathbf{u}^k) + \mathbf{y}_{\diamondsuit}^{k+1}/\varrho\|^2 \\ &= \underset{\mathbf{u}\in\mathbf{dom}\,g}{\arg\min}\; g(\mathbf{u}) + \langle \mathbf{v}^k + \mathbf{A}^T\mathbf{y}_{\diamondsuit}^{k+1}, \mathbf{u}-\mathbf{u}^k\rangle + \frac{r_k}{2\eta_k}\langle \mathbf{u}-\mathbf{u}^k, (\mathbf{W}^k + (\eta_k/r_k)\varrho\mathbf{A}^T\mathbf{A})(\mathbf{u}-\mathbf{u}^k)\rangle \\ &\overset{\text{(b)}}{=} \underset{\mathbf{u}\in\mathbf{dom}\,g}{\arg\min}\; g(\mathbf{u}) + \langle \mathbf{v}^k + \mathbf{A}^T\mathbf{y}_{\diamondsuit}^{k+1}, \mathbf{u}-\mathbf{u}^k\rangle + \frac{1}{2\widetilde{\eta}_k}\|\mathbf{u}-\mathbf{u}^k\|^2 \\ &= \underset{\mathbf{u}\in\mathbf{dom}\,g}{\arg\min}\; g(\mathbf{u}) + \frac{1}{2\widetilde{\eta}_k}\|\mathbf{u}-\mathbf{u}^k + \widetilde{\eta}_k(\mathbf{v}^k + \mathbf{A}^T\mathbf{y}_{\diamondsuit}^{k+1})\|^2 \\ &= \mathbf{prox}_{\widetilde{\eta}_k g}(\mathbf{u}^k - \widetilde{\eta}_k(\mathbf{v}^k + \mathbf{A}^T\mathbf{y}_{\diamondsuit}^{k+1})),\end{aligned} \tag{S-39}$$

where in (a) we use (S-38) and in (b) we use the definition of $\mathbf{W}^k$ in Section 4.2.

References

[1] A. Nemirovski, A. Juditsky, G. Lan, and A. Shapiro, "Robust stochastic approximation approach to stochastic programming," *SIAM J. Optim.*, vol. 19, no. 4, pp. 1574–1609, 2009.

[2] H. Ouyang, N. He, L. Tran, and A. Gray, "Stochastic alternating direction method of multipliers," in *Proc. ICML*, (Atlanta, USA), pp. 80–88, 2013.

[3] T. Suzuki, "Dual averaging and proximal gradient descent for online alternating direction multiplier method," in *Proc. ICML*, pp. 392–400, 2013.

[4] S. Azadi and S. Sra, "Towards an optimal stochastic alternating direction method of multipliers," in *Proc. ICML*, (Beijing, China), pp. 620–628, 2014.

[5] E. Esser, X. Zhang, and T. F. Chan, "A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science," *SIAM J. Imaging Sci.*, vol. 3, no. 4, pp. 1015–1046, 2010.

[6] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, pp. 1–122, Jan. 2011.

[7] M. Yan and W. Yin, *Self Equivalence of the Alternating Direction Method of Multipliers*, pp. 165–194. Springer International Publishing, 2016.