Supplemental Material to "Stochastic Three-Composite Convex Minimization with a Linear Operator"

Renbo Zhao* Volkan Cevher†

S-1 Proof of Theorem 1

From (5) and (6), we have

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

$$\frac{1}{\tau_k}(\mathbf{x}^k - \mathbf{x}^{k+1}) - (\mathbf{v}^k + \mathbf{A}^T\mathbf{y}^{k+1}) \in \partial g(\mathbf{x}^{k+1}) \tag{S-2}$$

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

$$\begin{aligned} h^*(\mathbf{y}) &\geq h^*(\mathbf{y}^{k+1}) + \left\langle \frac{1}{\alpha_k}(\mathbf{y}^k - \mathbf{y}^{k+1}) + \mathbf{A}\mathbf{z}^k, \mathbf{y} - \mathbf{y}^{k+1} \right\rangle \\ &= h^*(\mathbf{y}^{k+1}) + \frac{1}{\alpha_k}\langle \mathbf{y}^k - \mathbf{y}^{k+1}, \mathbf{y} - \mathbf{y}^{k+1}\rangle + \langle \mathbf{A}\mathbf{z}^k, \mathbf{y} - \mathbf{y}^{k+1}\rangle, \\ &= h^*(\mathbf{y}^{k+1}) + \frac{1}{2\alpha_k}\left(\|\mathbf{y}^k - \mathbf{y}^{k+1}\|^2 + \|\mathbf{y}^{k+1} - \mathbf{y}\|^2 - \|\mathbf{y}^k - \mathbf{y}\|^2\right) \\ &\qquad + \langle \mathbf{A}(\mathbf{x}^k + \theta_k(\mathbf{x}^k - \mathbf{x}^{k-1})), \mathbf{y} - \mathbf{y}^{k+1}\rangle, \end{aligned} \tag{S-3}$$

$$\begin{aligned} g(\mathbf{x}) &\geq g(\mathbf{x}^{k+1}) + \left\langle \frac{1}{\tau_k}(\mathbf{x}^k - \mathbf{x}^{k+1}) - (\mathbf{v}^k + \mathbf{A}^T\mathbf{y}^{k+1}), \mathbf{x} - \mathbf{x}^{k+1} \right\rangle \\ &= g(\mathbf{x}^{k+1}) + \frac{1}{\tau_k}\langle \mathbf{x}^k - \mathbf{x}^{k+1}, \mathbf{x} - \mathbf{x}^{k+1}\rangle - \langle \mathbf{v}^k + \mathbf{A}^T\mathbf{y}^{k+1}, \mathbf{x} - \mathbf{x}^{k+1}\rangle \\ &= g(\mathbf{x}^{k+1}) + \frac{1}{2\tau_k}\left(\|\mathbf{x}^k - \mathbf{x}^{k+1}\|^2 + \|\mathbf{x}^{k+1} - \mathbf{x}\|^2 - \|\mathbf{x}^k - \mathbf{x}\|^2\right) \\ &\qquad - \langle \mathbf{v}^k, \mathbf{x} - \mathbf{x}^{k+1}\rangle - \langle \mathbf{A}^T\mathbf{y}^{k+1}, \mathbf{x} - \mathbf{x}^{k+1}\rangle. \end{aligned} \tag{S-4}$$

Since f is convex and L-smooth, we have for any $\mathbf{x} \in \mathbb{R}^d$,

$$\begin{aligned} f(\mathbf{x}) &\geq f(\mathbf{x}^k) + \langle \nabla f(\mathbf{x}^k), \mathbf{x} - \mathbf{x}^k\rangle \\ &\geq f(\mathbf{x}^{k+1}) - \langle \nabla f(\mathbf{x}^k), \mathbf{x}^{k+1} - \mathbf{x}^k\rangle - \frac{L}{2}\|\mathbf{x}^k - \mathbf{x}^{k+1}\|^2 + \langle \nabla f(\mathbf{x}^k), \mathbf{x} - \mathbf{x}^k\rangle \\ &= f(\mathbf{x}^{k+1}) + \langle \nabla f(\mathbf{x}^k), \mathbf{x} - \mathbf{x}^{k+1}\rangle - \frac{L}{2}\|\mathbf{x}^k - \mathbf{x}^{k+1}\|^2. \end{aligned} \tag{S-5}$$

*Laboratory for Information and Inference Systems (LIONS), École Polytechnique Fédérale de Lausanne (EPFL), CH1015 - Lausanne, Switzerland. Email: **elezren@nus.edu.sg**.

†Laboratory for Information and Inference Systems (LIONS), École Polytechnique Fédérale de Lausanne (EPFL), CH1015 - Lausanne, Switzerland. Email: **volkan.cevher@epfl.ch**.

Summing (S-3), (S-4) and (S-5) and recalling $\boldsymbol{\varepsilon}^k = \nabla f(\mathbf{x}^k) - \mathbf{v}^k$, we have

$$
\begin{aligned}
0 \geq\ & (f(\mathbf{x}^{k+1}) - f(\mathbf{x})) + (g(\mathbf{x}^{k+1}) - g(\mathbf{x})) + (h^*(\mathbf{y}^{k+1}) - h^*(\mathbf{y})). \\
& + \left(\frac{1}{2\tau_k} - \frac{L}{2}\right)\|\mathbf{x}^k - \mathbf{x}^{k+1}\|^2 + \theta_k\langle \mathbf{A}(\mathbf{x}^k - \mathbf{x}^{k-1}), \mathbf{y}^k - \mathbf{y}^{k+1}\rangle \\
& + \frac{1}{2\tau_k}\left(\|\mathbf{x}^{k+1} - \mathbf{x}\|^2 - \|\mathbf{x}^k - \mathbf{x}\|^2\right) + \frac{1}{2\alpha_k}\left(\|\mathbf{y}^{k+1} - \mathbf{y}\|^2 - \|\mathbf{y}^k - \mathbf{y}\|^2\right) \\
& - \langle \mathbf{A}(\mathbf{x}^{k+1} - \mathbf{x}^k), \mathbf{y} - \mathbf{y}^{k+1}\rangle + \theta_k\langle \mathbf{A}(\mathbf{x}^k - \mathbf{x}^{k-1}), \mathbf{y} - \mathbf{y}^k\rangle \\
& + \frac{1}{2\alpha_k}\|\mathbf{y}^k - \mathbf{y}^{k+1}\|^2 + \left(\langle \mathbf{A}\mathbf{x}^{k+1}, \mathbf{y}\rangle - \langle \mathbf{A}\mathbf{x}, \mathbf{y}^{k+1}\rangle\right) + \langle \boldsymbol{\varepsilon}^k, \mathbf{x} - \mathbf{x}^{k+1}\rangle.
\end{aligned} \tag{S-6}
$$

Using Cauchy-Schwartz and Young's inequality, we have for any $\mu_k > 0$,

$$
\begin{aligned}
\langle \mathbf{A}(\mathbf{x}^k - \mathbf{x}^{k-1}), \mathbf{y}^k - \mathbf{y}^{k+1}\rangle &\geq -B\|\mathbf{x}^k - \mathbf{x}^{k-1}\|\|\mathbf{y}^{k+1} - \mathbf{y}^k\| \\
&\geq -\frac{\mu_k B}{2}\|\mathbf{x}^k - \mathbf{x}^{k-1}\|^2 - \frac{B}{2\mu_k}\|\mathbf{y}^{k+1} - \mathbf{y}^k\|^2.
\end{aligned} \tag{S-7}
$$

Define $\widetilde{\mathbf{x}}^{k+1} \triangleq \mathbf{prox}_{\tau_k g}(\mathbf{x}^k - \tau_k(\mathbf{A}^T\mathbf{y}^{k+1} + \nabla f(\mathbf{x}^k)))$. Using the nonexpansiveness of $\mathbf{prox}_{\tau_k g}$ in (6), we have

$$
\|\widetilde{\mathbf{x}}^{k+1} - \mathbf{x}^{k+1}\| \leq \tau_k\|\boldsymbol{\varepsilon}^k\|. \tag{S-8}
$$

By (S-8) and Cauchy-Schwartz, we have

$$
\begin{aligned}
\langle \boldsymbol{\varepsilon}^k, \mathbf{x} - \mathbf{x}^{k+1}\rangle &= \langle \boldsymbol{\varepsilon}^k, \mathbf{x} - \widetilde{\mathbf{x}}^{k+1}\rangle + \langle \boldsymbol{\varepsilon}^k, \widetilde{\mathbf{x}}^{k+1} - \mathbf{x}^{k+1}\rangle \\
&\geq \langle \boldsymbol{\varepsilon}^k, \mathbf{x} - \widetilde{\mathbf{x}}^{k+1}\rangle - \tau_k\|\boldsymbol{\varepsilon}^k\|^2.
\end{aligned} \tag{S-9}
$$

Now, substitute (S-7) and (S-9) into (S-6) and then multiply both sides of (S-6) by τ_k, we have

$$
\begin{aligned}
& -\tau_k\theta_k\langle \mathbf{A}(\mathbf{x}^k - \mathbf{x}^{k-1}), \mathbf{y} - \mathbf{y}^k\rangle + \frac{\tau_k\theta_k\mu_k B}{2}\|\mathbf{x}^k - \mathbf{x}^{k-1}\|^2 + \frac{1}{2}\|\mathbf{x}^k - \mathbf{x}\|^2 \\
& + \frac{\tau_k}{2\alpha_k}\|\mathbf{y}^k - \mathbf{y}\|^2 \geq \tau_k\left(L\left(\mathbf{x}^{k+1}, \mathbf{y}\right) - L\left(\mathbf{x}, \mathbf{y}^{k+1}\right)\right) - \tau_k\langle \mathbf{A}(\mathbf{x}^{k+1} - \mathbf{x}^k), \mathbf{y} - \mathbf{y}^{k+1}\rangle \\
& + \left(\frac{1}{2} - \frac{\tau_k L}{2}\right)\|\mathbf{x}^k - \mathbf{x}^{k+1}\|^2 + \tau_k\left(\frac{1}{2\alpha_k} - \frac{\theta_k B}{2\mu_k}\right)\|\mathbf{y}^k - \mathbf{y}^{k+1}\|^2 \\
& + \frac{1}{2}\|\mathbf{x}^{k+1} - \mathbf{x}\|^2 + \frac{\tau_k}{2\alpha_k}\|\mathbf{y}^{k+1} - \mathbf{y}\|^2 + \tau_k\langle \boldsymbol{\varepsilon}^k, \mathbf{x} - \widetilde{\mathbf{x}}^{k+1}\rangle - \tau_k^2\|\boldsymbol{\varepsilon}^k\|^2.
\end{aligned} \tag{S-10}
$$

By choosing $\mu_k = \theta_k\alpha_k B$ and using conditions (13) and (14), we have

$$
\begin{aligned}
& -\tau_k\theta_k\langle \mathbf{A}(\mathbf{x}^k - \mathbf{x}^{k-1}), \mathbf{y} - \mathbf{y}^k\rangle + \frac{\tau_k\theta_k^2\alpha_k B^2}{2}\|\mathbf{x}^k - \mathbf{x}^{k-1}\|^2 + \frac{1}{2}\|\mathbf{x}^k - \mathbf{x}\|^2 \\
& + \frac{\tau_k}{2\alpha_k}\|\mathbf{y}^k - \mathbf{y}\|^2 \geq \tau_k\left(L\left(\mathbf{x}^{k+1}, \mathbf{y}\right) - L\left(\mathbf{x}, \mathbf{y}^{k+1}\right)\right) \\
& - \tau_{k+1}\theta_{k+1}\langle \mathbf{A}(\mathbf{x}^{k+1} - \mathbf{x}^k), \mathbf{y} - \mathbf{y}^{k+1}\rangle + \frac{\tau_{k+1}\theta_{k+1}^2\alpha_{k+1}B^2}{2}\|\mathbf{x}^k - \mathbf{x}^{k+1}\|^2 \\
& + \frac{1}{2}\|\mathbf{x}^{k+1} - \mathbf{x}\|^2 + \frac{\tau_k}{2\alpha_k}\|\mathbf{y}^{k+1} - \mathbf{y}\|^2 + \tau_k\langle \boldsymbol{\varepsilon}^k, \mathbf{x} - \widetilde{\mathbf{x}}^{k+1}\rangle - \tau_k^2\|\boldsymbol{\varepsilon}^k\|^2.
\end{aligned} \tag{S-11}
$$

For any $k \geq 1$, define $q_k \triangleq \tau_k/\alpha_k = \tau_{k-1}^2 B^2/(1 - \tau_{k-1}L)$. We claim that $\{q_k\}_{k\geq 1}$ is a (strictly) decreasing sequence. Indeed, from (12), we see that $\{\tau_k\}_{k\geq 0}$ is strictly decreasing. Then

$$
\frac{q_k}{q_{k+1}} = \frac{\tau_{k-1}^2 B^2}{1 - \tau_{k-1}L} \cdot \frac{1 - \tau_k L}{\tau_k^2 B^2} = \frac{\tau_{k-1}}{\tau_k} \cdot \frac{\tau_{k-1} - \tau_{k-1}\tau_k L}{\tau_k - \tau_{k-1}\tau_k L} > 1.
$$

Moreover, since $q_0 > q_1$ (by the choice of α_0), $\{q_k\}_{k\geq 0}$ is (strictly) decreasing. Based on this fact, we take supremum on both sides of (S-11) over any bounded sets $\mathcal{X}' \subseteq \mathbb{R}^d$ and $\mathcal{Y}' \subseteq \mathbb{R}^m$ to obtain

$$\begin{aligned} &-\inf_{\mathbf{y}\in\mathcal{Y}'} \tau_k\theta_k\langle \mathbf{A}(\mathbf{x}^k-\mathbf{x}^{k-1}), \mathbf{y}-\mathbf{y}^k\rangle + \frac{\tau_k\theta_k^2\alpha_k B^2}{2}\|\mathbf{x}^k-\mathbf{x}^{k-1}\|^2 \\ &+\sup_{\mathbf{x}\in\mathcal{X}'}\frac{1}{2}\|\mathbf{x}^k-\mathbf{x}\|^2 + \sup_{\mathbf{y}\in\mathcal{Y}'}\frac{\tau_k}{2\alpha_k}\|\mathbf{y}^k-\mathbf{y}\|^2 \geq \tau_k G_{\mathcal{X}',\mathcal{Y}'}(\mathbf{x}^{k+1},\mathbf{y}^{k+1}) \\ &-\inf_{\mathbf{y}\in\mathcal{Y}'}\tau_{k+1}\theta_{k+1}\langle \mathbf{A}(\mathbf{x}^{k+1}-\mathbf{x}^k),\mathbf{y}-\mathbf{y}^{k+1}\rangle + \frac{\tau_{k+1}\theta_{k+1}^2\alpha_{k+1}B^2}{2}\|\mathbf{x}^k-\mathbf{x}^{k+1}\|^2 \\ &+\sup_{\mathbf{x}\in\mathcal{X}'}\frac{1}{2}\|\mathbf{x}^{k+1}-\mathbf{x}\|^2 + \sup_{\mathbf{y}\in\mathcal{Y}'}\frac{\tau_{k+1}}{2\alpha_{k+1}}\|\mathbf{y}^{k+1}-\mathbf{y}\|^2 + \sup_{\mathbf{x}\in\mathcal{X}'}\tau_k\langle\boldsymbol{\varepsilon}^k,\mathbf{x}-\widetilde{\mathbf{x}}^{k+1}\rangle - \tau_k^2\|\boldsymbol{\varepsilon}^k\|^2. \end{aligned} \tag{S-12}$$

Using $\mathbb{E}_{\boldsymbol{\xi}^k}\left[\mathbf{v}^k \mid \mathcal{F}_k\right] = \nabla f(\mathbf{x}^k)$ in Assumption 1 and reverse Fatou's lemma [1], we have

$$\mathbb{E}_{\boldsymbol{\xi}^k}\left[\sup_{\mathbf{x}\in\mathcal{X}'}\tau_k\langle\boldsymbol{\varepsilon}^k,\mathbf{x}-\widetilde{\mathbf{x}}^{k+1}\rangle \,\middle|\, \mathcal{F}_k\right] \geq \sup_{\mathbf{x}\in\mathcal{X}'}\tau_k\left\langle \mathbb{E}_{\boldsymbol{\xi}^k}[\boldsymbol{\varepsilon}^k\mid\mathcal{F}_k],\mathbf{x}-\widetilde{\mathbf{x}}^{k+1}\right\rangle = 0. \tag{S-13}$$

Based on this, we take expectation on the RHS of (S-12) w.r.t. $\boldsymbol{\xi}_k$ by conditioning on $\mathcal{F}_k$,

$$\begin{aligned} &-\inf_{\mathbf{y}\in\mathcal{Y}'} \tau_k\theta_k\langle \mathbf{A}(\mathbf{x}^k-\mathbf{x}^{k-1}), \mathbf{y}-\mathbf{y}^k\rangle + \frac{\tau_k\theta_k^2\alpha_k B^2}{2}\|\mathbf{x}^k-\mathbf{x}^{k-1}\|^2 \\ &+\sup_{\mathbf{x}\in\mathcal{X}'}\frac{1}{2}\|\mathbf{x}^k-\mathbf{x}\|^2 + \sup_{\mathbf{y}\in\mathcal{Y}'}\frac{\tau_k}{2\alpha_k}\|\mathbf{y}^k-\mathbf{y}\|^2 \geq \mathbb{E}_{\boldsymbol{\xi}^k}\Big[\tau_k G_{\mathcal{X}',\mathcal{Y}'}(\mathbf{x}^{k+1},\mathbf{y}^{k+1}) \\ &-\inf_{\mathbf{y}\in\mathcal{Y}'}\tau_{k+1}\theta_{k+1}\langle \mathbf{A}(\mathbf{x}^{k+1}-\mathbf{x}^k),\mathbf{y}-\mathbf{y}^{k+1}\rangle + \frac{\tau_{k+1}\theta_{k+1}^2\alpha_{k+1}B^2}{2}\|\mathbf{x}^k-\mathbf{x}^{k+1}\|^2 \\ &+\sup_{\mathbf{x}\in\mathcal{X}'}\frac{1}{2}\|\mathbf{x}^{k+1}-\mathbf{x}\|^2 + \sup_{\mathbf{y}\in\mathcal{Y}'}\frac{\tau_{k+1}}{2\alpha_{k+1}}\|\mathbf{y}^{k+1}-\mathbf{y}\|^2 \,\Big|\, \mathcal{F}_k\Big] - \tau_k^2\mathbb{E}_{\boldsymbol{\xi}^k}\left[\|\boldsymbol{\varepsilon}^k\|^2 \,\middle|\, \mathcal{F}_k\right]. \end{aligned} \tag{S-14}$$

Telescoping (S-14) over $k = 0, 1, \ldots, K-1$,

$$\begin{aligned} &-\inf_{\mathbf{y}\in\mathcal{Y}'}\tau_0\theta_0\langle \mathbf{A}(\mathbf{x}^0-\mathbf{x}^{-1}),\mathbf{y}-\mathbf{y}^0\rangle + \frac{\tau_0\theta_0^2\alpha_0B^2}{2}\|\mathbf{x}^0-\mathbf{x}^{-1}\|^2 \\ &+\sup_{\mathbf{x}\in\mathcal{X}'}\frac{1}{2}\|\mathbf{x}^0-\mathbf{x}\|^2 + \sup_{\mathbf{y}\in\mathcal{Y}'}\frac{\tau_0}{2\alpha_0}\|\mathbf{y}^0-\mathbf{y}\|^2 \geq \mathbb{E}_{\Xi_k}\left[\sum_{k=0}^{K-1}\tau_k G_{\mathcal{X}',\mathcal{Y}'}(\mathbf{x}^{k+1},\mathbf{y}^{k+1})\right] \\ &+\mathbb{E}_{\Xi_k}\Big[-\inf_{\mathbf{y}\in\mathcal{Y}'}\tau_K\theta_K\langle \mathbf{A}(\mathbf{x}^K-\mathbf{x}^{K-1}),\mathbf{y}-\mathbf{y}^K\rangle + \sup_{\mathbf{x}\in\mathcal{X}'}\frac{1}{2}\|\mathbf{x}^K-\mathbf{x}\|^2 \\ &+\frac{\tau_K\theta_K^2\alpha_KB^2}{2}\|\mathbf{x}^K-\mathbf{x}^{K-1}\|^2 + \sup_{\mathbf{y}\in\mathcal{Y}'}\frac{\tau_K}{2\alpha_K}\|\mathbf{y}^K-\mathbf{y}\|^2\Big] - \sum_{k=0}^{K-1}\tau_k^2\mathbb{E}_{\boldsymbol{\xi}^k}\left[\|\boldsymbol{\varepsilon}^k\|^2 \,\middle|\, \mathcal{F}_k\right]. \end{aligned} \tag{S-15}$$

Since $\mathbf{z}^0 = \mathbf{x}^0$, we have $\mathbf{x}^0 = \mathbf{x}^{-1}$. By Young's inequality, we have

$$\begin{aligned} &\inf_{\mathbf{y}\in\mathcal{Y}'}\langle \tau_K\theta_K\mathbf{A}(\mathbf{x}^K-\mathbf{x}^{K-1}),\mathbf{y}-\mathbf{y}^K\rangle \\ &\leq \frac{\alpha_K\tau_K\theta_K^2B^2}{2}\|\mathbf{x}^K-\mathbf{x}^{K-1}\|^2 + \inf_{\mathbf{y}\in\mathcal{Y}'}\frac{\tau_K}{2\alpha_K}\|\mathbf{y}-\mathbf{y}^K\|^2 \\ &\leq \frac{\alpha_K\tau_K\theta_K^2B^2}{2}\|\mathbf{x}^K-\mathbf{x}^{K-1}\|^2 + \sup_{\mathbf{y}\in\mathcal{Y}'}\frac{\tau_K}{2\alpha_K}\|\mathbf{y}-\mathbf{y}^K\|^2. \end{aligned} \tag{S-16}$$

Since we use option I in Algorithm 1, we have

$$S_K = \sum_{k=0}^{K-1}\tau_k, \quad \overline{\mathbf{x}}^K = \frac{1}{S_K}\sum_{k=1}^{K}\tau_{k-1}\mathbf{x}^k, \quad \overline{\mathbf{y}}^K = \frac{1}{S_K}\sum_{k=1}^{K}\tau_{k-1}\mathbf{y}^k.$$

From the joint convexity of $G_{\mathcal{X}',\mathcal{Y}'}(\cdot,\cdot)$, we employ Jensen's inequality to obtain

$$S_K \frac{1}{S_K} \sum_{k=0}^{K-1} \tau_k G_{\mathcal{X}',\mathcal{Y}'}(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}) \geq S_K G_{\mathcal{X}',\mathcal{Y}'}(\overline{\mathbf{x}}^K, \overline{\mathbf{y}}^K). \tag{S-17}$$

Based on (S-16), (S-17) and $\mathbf{x}^{-1} = \mathbf{x}^0$, we have

$$S_K \mathbb{E}_{\Xi_K}\left[G_{\mathcal{X}',\mathcal{Y}'}(\overline{\mathbf{x}}^K, \overline{\mathbf{y}}^K)\right] \leq \frac{1}{2} R^2_{\mathcal{X}'}(\mathbf{x}^0) + \frac{\tau_0}{2\alpha_0} R^2_{\mathcal{Y}'}(\mathbf{y}^0) + \sum_{k=0}^{K-1} \tau_k^2 \mathbb{E}_{\boldsymbol{\xi}^k}\left[\|\boldsymbol{\varepsilon}^k\|^2 \,\middle|\, \mathcal{F}_k\right].$$

Define $\widetilde{S}_K \triangleq \sum_{k=0}^{K-1} \tau_k^2$. Using $\mathbb{E}_{\boldsymbol{\xi}^k}\left[\|\boldsymbol{\varepsilon}^k\|^2 \,|\, \mathcal{F}_k\right] \leq \sigma^2$ in Assumption 1, we have

$$\mathbb{E}_{\Xi_K}\left[G_{\mathcal{X}',\mathcal{Y}'}(\overline{\mathbf{x}}^K, \overline{\mathbf{y}}^K)\right] \leq \frac{1}{2S_K}\left(R^2_{\mathcal{X}'}(\mathbf{x}^0) + \frac{\tau_0}{\alpha_0} R^2_{\mathcal{Y}'}(\mathbf{y}^0)\right) + \frac{\widetilde{S}_K}{S_K}\sigma^2. \tag{S-18}$$

S-2 Proof of Corollary 1

Since $S_K = K\tau_K$ and $\widetilde{S}_K = K\tau_K^2$, from (20), we have

$$\begin{aligned}
\mathbb{E}_{\Xi_K}\left[G(\overline{\mathbf{x}}^K, \overline{\mathbf{y}}^K)\right] &\leq \frac{1}{2K\tau_K}\left(R_g^2(\mathbf{x}^0) + \frac{\tau_K}{\alpha_K} R^2_{h^*}(\mathbf{y}^0)\right) + \tau_K\sigma^2 \\
&\overset{\text{(a)}}{=} \frac{R_g^2(\mathbf{x}^0)}{2K\tau_K} + \tau_K\left(\frac{B^2 R^2_{h^*}(\mathbf{y}^0)}{2K(1-L\tau_K)} + \sigma^2\right) \\
&\overset{\text{(b)}}{\leq} \frac{R_g^2(\mathbf{x}^0)}{2K\tau_K} + \tau_K\left(\frac{B^2 R^2_{h^*}(\mathbf{y}^0)}{2K(1-\widetilde{r})} + \sigma^2\right),
\end{aligned} \tag{S-19}$$

where (a) follows from the choice of α_K and (b) follows from $\tau_K \leq \widetilde{r}/L$. For convenience, for any $K \geq 1$, define

$$E_K \triangleq \frac{B^2 R^2_{h^*}(\mathbf{y}^0)}{2K(1-\widetilde{r})} + \frac{3}{2}\sigma^2. \tag{S-20}$$

If we choose $\widetilde{a}$, $\widetilde{b}$ and $\widetilde{b}'$ as in (21), then (11) becomes

$$\tau_K = \min\left\{\frac{\widetilde{r}}{L}, \frac{R_g(\mathbf{x}^0)}{\sqrt{2KE_K}}\right\} \tag{S-21}$$

and (22) becomes $L \geq \widetilde{r}\sqrt{2KE_K}/R_g(\mathbf{x}^0)$.

Now let us consider two cases, depending on the value of L.

Case I: Condition (22) holds. In this case, $\tau_K = \widetilde{r}/L$. Based on (S-19),

$$\begin{aligned}
\mathbb{E}_{\Xi_K}\left[G(\overline{\mathbf{x}}^K, \overline{\mathbf{y}}^K)\right] &\leq \frac{R_g^2(\mathbf{x}^0)L}{2K\widetilde{r}} + \frac{\widetilde{r}}{L}E_K \\
&\overset{\text{(a)}}{\leq} \frac{R_g^2(\mathbf{x}^0)L}{2K\widetilde{r}} + \frac{R_g(\mathbf{x}^0)\sqrt{E_K}}{\sqrt{2K}} \\
&\overset{\text{(b)}}{\leq} \frac{R_g^2(\mathbf{x}^0)L}{2K\widetilde{r}} + \frac{R_g(\mathbf{x}^0)\left(BR_{h^*}(\mathbf{y}^0)/\sqrt{2K(1-\widetilde{r})} + \sqrt{3/2}\sigma\right)}{\sqrt{2K}} \\
&= \frac{R_g^2(\mathbf{x}^0)L}{2K\widetilde{r}} + \frac{R_g(\mathbf{x}^0)R_{h^*}(\mathbf{y}^0)B}{2K\sqrt{1-\widetilde{r}}} + \frac{\sqrt{3}R_g(\mathbf{x}^0)\sigma}{2\sqrt{K}}
\end{aligned} \tag{S-22}$$

where (a) follows from (22) and (b) follows from $\sqrt{x+y} \leq \sqrt{x} + \sqrt{y}$ for any $x, y \geq 0$.

Case II: Condition (22) does not hold. Then $\tau_K = R_g(\mathbf{x}^0)/\sqrt{2KE_K}$. From (S-19),

$$\begin{aligned}\mathbb{E}_{\Xi_K}\left[G(\overline{\mathbf{x}}^K, \overline{\mathbf{y}}^K)\right] \leq \frac{\sqrt{2E_K}R_g(\mathbf{x}^0)}{\sqrt{K}} &\leq \frac{\sqrt{2}R_g(\mathbf{x}^0)\left(BR_{h^*}(\mathbf{y}^0)/\sqrt{2K(1-\widetilde{r})} + \sqrt{3/2}\sigma\right)}{\sqrt{K}} \\ &= \frac{R_g(\mathbf{x}^0)R_{h^*}(\mathbf{y}^0)B}{K\sqrt{1-\widetilde{r}}} + \frac{\sqrt{3}R_g(\mathbf{x}^0)\sigma}{\sqrt{K}} \end{aligned} \tag{S-23}$$

S-3 Proof of Corollary 2

In (12), if we choose $a = \widetilde{a}$, $b = \widetilde{b}$, $b' = \widetilde{b}' + 1$ and $r = \widetilde{r}$, then for any $0 \leq k \leq K-1$,

$$\tau_k = \min\left\{\frac{\widetilde{r}}{L}, \frac{R_g(\mathbf{x}^0)}{\sqrt{2(k+1)E_{k+1}}}\right\}. \tag{S-24}$$

Similar to the proof of Corollary 1 (shown in Appendix S-2), we consider two cases.

Case I: Condition (22) holds. This case coincides with Case I of Appendix S-2.

Case II: Condition (22) does not hold. In this case, let K' be the largest integer from 0 to $K-2$ such that $\tau_{K'} = \widetilde{r}/L$, i.e.,

$$K' \triangleq \left\lfloor \frac{1}{3\sigma^2}\left(\frac{L^2R_g^2(\mathbf{x}^0)}{\widetilde{r}^2} - \frac{B^2R_{h^*}^2(\mathbf{y}^0)}{1-\widetilde{r}}\right)\right\rfloor. \tag{S-25}$$

As a result,

$$\begin{aligned} S_K &= \sum_{k=0}^{K'}\frac{\widetilde{r}}{L} + \sum_{k=K'+1}^{K-1}\frac{R_g(\mathbf{x}^0)}{\sqrt{B^2R_{h^*}^2(\mathbf{y}^0)/(1-\widetilde{r}) + 3\sigma^2k}} \\ &\geq (K'+1)\frac{\widetilde{r}}{L} + \int_{K'+1}^{K}\frac{R_g(\mathbf{x}^0)}{\sqrt{B^2R_{h^*}^2(\mathbf{y}^0)/(1-\widetilde{r}) + 3\sigma^2z}}\mathrm{d}z \\ &\geq (K'+1)\frac{\widetilde{r}}{L} + R_g(\mathbf{x}^0)\frac{K-(K'+1)}{\sqrt{B^2R_{h^*}^2(\mathbf{y}^0)/(1-\widetilde{r}) + 3\sigma^2K}} \\ &\geq (K'+1)\left(\frac{\widetilde{r}}{L} - \frac{R_g(\mathbf{x}^0)}{\sqrt{B^2R_{h^*}^2(\mathbf{y}^0)/(1-\widetilde{r}) + 3\sigma^2K}}\right) + \frac{R_g(\mathbf{x}^0)K}{\sqrt{B^2R_{h^*}^2(\mathbf{y}^0)/(1-\widetilde{r}) + 3\sigma^2K}} \\ &\geq \frac{R_g(\mathbf{x}^0)K}{\sqrt{B^2R_{h^*}^2(\mathbf{y}^0)/(1-\widetilde{r}) + 3\sigma^2K}}. \end{aligned} \tag{S-26}$$

In addition,

$$\begin{aligned} \widetilde{S}_K &= \sum_{k=0}^{K'}\frac{\widetilde{r}^2}{L^2} + \sum_{k=K'+1}^{K-1}\frac{R_g^2(\mathbf{x}^0)}{B^2R_{h^*}^2(\mathbf{y}^0)/(1-\widetilde{r}) + 3\sigma^2k} \\ &\overset{\text{(a)}}{\leq} \sum_{k=0}^{K-1}\frac{R_g^2(\mathbf{x}^0)}{B^2R_{h^*}^2(\mathbf{y}^0)/(1-\widetilde{r}) + 3\sigma^2k} \\ &\leq \frac{R_g^2(\mathbf{x}^0)(1-\widetilde{r})}{B^2R_{h^*}^2(\mathbf{y}^0)} + \int_0^{K-1}\frac{R_g^2(\mathbf{x}^0)}{B^2R_{h^*}^2(\mathbf{y}^0)/(1-\widetilde{r}) + 3\sigma^2z}\mathrm{d}z \\ &\leq \frac{R_g^2(\mathbf{x}^0)(1-\widetilde{r})}{B^2R_{h^*}^2(\mathbf{y}^0)} - \frac{R_g^2(\mathbf{x}^0)}{3\sigma^2}\log\left(\frac{B^2R_{h^*}^2(\mathbf{y}^0)}{1-\widetilde{r}}\right) + \frac{R_g^2(\mathbf{x}^0)}{3\sigma^2}\log\left(\frac{B^2R_{h^*}^2(\mathbf{y}^0)}{1-\widetilde{r}} + 3\sigma^2K\right) \end{aligned}$$

$$\stackrel{\text{(b)}}{\leq} \left(1+\frac{1}{3\sigma^2}\right)\frac{R_g^2(\mathbf{x}^0)(1-\widetilde{r})}{B^2R_{h^*}^2(\mathbf{y}^0)} + \frac{R_g^2(\mathbf{x}^0)}{3\sigma^2}\log\left(\frac{B^2R_{h^*}^2(\mathbf{y}^0)}{1-\widetilde{r}} + 3\sigma^2 K\right) = C_K, \tag{S-27}$$

where in (a) we use the definition of K' in (S-25) and in (b) we use $1-1/x \leq \log x$ for all $x > 0$. Since $\tau_0/\alpha_0 \leq 2\tau_1/\alpha_1$ and $\tau_0 \leq \min\left\{\frac{\widetilde{r}}{L}, \frac{R_g(\mathbf{x}^0)\sqrt{1-\widetilde{r}}}{BR_{h^*}(\mathbf{y}^0)}\right\}$, we also have

$$\frac{\tau_0}{\alpha_0} \leq 2\frac{\tau_0^2 B^2}{1-L\tau_0} \leq 2\frac{B^2}{1-\widetilde{r}}\frac{R_g^2(\mathbf{x}^0)(1-\widetilde{r})}{B^2R_{h^*}^2(\mathbf{y}^0)} = \frac{2R_g^2(\mathbf{x}^0)}{R_{h^*}^2(\mathbf{y}^0)}. \tag{S-28}$$

Using (S-26), (S-27) and (S-28), we obtain (25) from (20).

S-4 Proof of Theorem 2

We first present a lemma that will be used in our proof. See Appendix S-5 for the proof of this lemma.

Lemma S-1. *If the positive sequences $\{\alpha_k\}_{k\geq 0}$, $\{\tau_k\}_{k\geq 0}$ and $\{\theta_k\}_{k\geq 0}$ satisfy $\alpha_0 \geq 1$, $\theta_0 > 0$ and conditions* (15)*,* (16) *and* (17)*, then for any $k \geq 1$,*

$$\alpha_0 + \frac{\gamma}{2B^2+2L+\gamma}k \leq \alpha_k \leq \alpha_0 + \frac{\gamma}{B^2}k, \tag{S-29}$$

$$\frac{1}{B^2\alpha_0 + L + \gamma k} \leq \tau_k \leq \frac{2B^2+2L+\gamma}{(2B^2+2L+\gamma)(B^2\alpha_0+L)+B^2\gamma k}. \tag{S-30}$$

In particular, we have $\alpha_k = \Theta(k)$ and $\tau_k = \Theta(1/k)$. Consequently, from (15)*, we have $\theta_k = \Theta(1)$.*

S-4.1 Proof of Part (i)

By the strong convexity of g, we have for any $\mathbf{x} \in \mathbb{R}^d$,

$$\begin{aligned} g(\mathbf{x}) &\geq g(\mathbf{x}^{k+1}) + \left\langle \frac{1}{\tau_k}(\mathbf{x}^k - \mathbf{x}^{k+1}) - (\mathbf{v}^k + \mathbf{A}^T\mathbf{y}^{k+1}), \mathbf{x} - \mathbf{x}^{k+1}\right\rangle + \frac{\gamma}{2}\|\mathbf{x}-\mathbf{x}^{k+1}\|^2 \\ &= g(\mathbf{x}^{k+1}) + \frac{1}{\tau_k}\langle \mathbf{x}^k - \mathbf{x}^{k+1}, \mathbf{x}-\mathbf{x}^{k+1}\rangle - \langle \mathbf{v}^k + \mathbf{A}^T\mathbf{y}^{k+1}, \mathbf{x}-\mathbf{x}^{k+1}\rangle + \frac{\gamma}{2}\|\mathbf{x}-\mathbf{x}^{k+1}\|^2 \\ &= g(\mathbf{x}^{k+1}) + \frac{1}{2\tau_k}\left(\|\mathbf{x}^k-\mathbf{x}^{k+1}\|^2 + \|\mathbf{x}^{k+1}-\mathbf{x}\|^2 - \|\mathbf{x}^k-\mathbf{x}\|^2\right) - \langle \mathbf{v}^k, \mathbf{x}-\mathbf{x}^{k+1}\rangle \\ &\qquad - \langle \mathbf{A}^T\mathbf{y}^{k+1}, \mathbf{x}-\mathbf{x}^{k+1}\rangle + \frac{\gamma}{2}\|\mathbf{x}-\mathbf{x}^{k+1}\|^2. \end{aligned} \tag{S-31}$$

Summing (S-3), (S-5) and (S-31), we have for any $\mathbf{x}\in\mathbb{R}^d$ and $\mathbf{y}\in\mathbb{R}^m$,

$$\begin{aligned} 0 &\geq (f(\mathbf{x}^{k+1}) - f(\mathbf{x})) + (g(\mathbf{x}^{k+1}) - g(\mathbf{x})) + (h^*(\mathbf{y}^{k+1}) - h^*(\mathbf{y})) + \frac{\gamma}{2}\|\mathbf{x}-\mathbf{x}^{k+1}\|^2 \\ &- \langle \mathbf{A}(\mathbf{x}^{k+1}-\mathbf{x}^k), \mathbf{y}-\mathbf{y}^{k+1}\rangle + \theta_k\langle \mathbf{A}(\mathbf{x}^k-\mathbf{x}^{k-1}), \mathbf{y}-\mathbf{y}^k\rangle \\ &+ \left(\frac{1}{2\tau_k} - \frac{L}{2}\right)\|\mathbf{x}^k-\mathbf{x}^{k+1}\|^2 + \frac{1}{2\alpha_k}\|\mathbf{y}^k-\mathbf{y}^{k+1}\|^2 + \theta_k\langle \mathbf{A}(\mathbf{x}^k-\mathbf{x}^{k-1}), \mathbf{y}^k-\mathbf{y}^{k+1}\rangle \\ &+ \frac{1}{2\tau_k}\left(\|\mathbf{x}^{k+1}-\mathbf{x}\|^2 - \|\mathbf{x}^k-\mathbf{x}\|^2\right) + \frac{1}{2\alpha_k}\left(\|\mathbf{y}^{k+1}-\mathbf{y}\|^2 - \|\mathbf{y}^k-\mathbf{y}\|^2\right) + \langle -\boldsymbol{\varepsilon}^k, \mathbf{x}-\mathbf{x}^{k+1}\rangle \\ &+ \left(\langle \mathbf{A}\mathbf{x}^{k+1}, \mathbf{y}\rangle - \langle \mathbf{A}\mathbf{x}, \mathbf{y}^{k+1}\rangle\right) - \langle \mathbf{A}(\mathbf{x}^{k+1}-\mathbf{x}^k), \mathbf{y}-\mathbf{y}^{k+1}\rangle + \theta_k\langle \mathbf{A}(\mathbf{x}^k-\mathbf{x}^{k-1}), \mathbf{y}-\mathbf{y}^k\rangle. \end{aligned} \tag{S-32}$$

Now, substitute (S-7) and (S-9) into (S-32) and rearrange, we have

$$
\begin{aligned}
& -\theta_k\langle \mathbf{A}(\mathbf{x}^k-\mathbf{x}^{k-1}),\mathbf{y}-\mathbf{y}^k\rangle+\frac{\theta_k\mu_k B}{2}\|\mathbf{x}^k-\mathbf{x}^{k-1}\|^2+\frac{1}{2\tau_k}\|\mathbf{x}^k-\mathbf{x}\|^2+\frac{1}{2\alpha_k}\|\mathbf{y}^k-\mathbf{y}\|^2 \\
\geq\ & \left(L\left(\mathbf{x}^{k+1},\mathbf{y}\right)-L\left(\mathbf{x},\mathbf{y}^{k+1}\right)\right)-\langle \mathbf{A}(\mathbf{x}^{k+1}-\mathbf{x}^k),\mathbf{y}-\mathbf{y}^{k+1}\rangle+\left(\frac{1}{2\tau_k}-\frac{L}{2}\right)\|\mathbf{x}^k-\mathbf{x}^{k+1}\|^2 \\
& +\left(\frac{1}{2\alpha_k}-\frac{\theta_k B}{2\mu_k}\right)\|\mathbf{y}^k-\mathbf{y}^{k+1}\|^2+\left(\frac{1}{2\tau_k}+\frac{\gamma}{2}\right)\|\mathbf{x}^{k+1}-\mathbf{x}\|^2+\frac{1}{2\alpha_k}\|\mathbf{y}^{k+1}-\mathbf{y}\|^2 \\
& -\langle \boldsymbol{\varepsilon}^k,\mathbf{x}-\widetilde{\mathbf{x}}^{k+1}\rangle-\tau_k\|\boldsymbol{\varepsilon}^k\|^2.
\end{aligned} \tag{S-33}
$$

Taking supremum on both sides of (S-33) over any bounded sets $\mathcal{X}'\subseteq\mathbb{R}^d$ and $\mathcal{Y}'\subseteq\mathbb{R}^m$, we have

$$
\begin{aligned}
& -\inf_{\mathbf{y}\in\mathcal{Y}'}\theta_k\langle \mathbf{A}(\mathbf{x}^k-\mathbf{x}^{k-1}),\mathbf{y}-\mathbf{y}^k\rangle+\frac{\theta_k\mu_k B}{2}\|\mathbf{x}^k-\mathbf{x}^{k-1}\|^2+\sup_{\mathbf{x}\in\mathcal{X}'}\frac{1}{2\tau_k}\|\mathbf{x}^k-\mathbf{x}\|^2 \\
& +\sup_{\mathbf{y}\in\mathcal{Y}'}\frac{1}{2\alpha_k}\|\mathbf{y}^k-\mathbf{y}\|^2\geq G_{\mathcal{X}',\mathcal{Y}'}(\mathbf{x}^{k+1},\mathbf{y}^{k+1})-\inf_{\mathbf{y}\in\mathcal{Y}'}\langle \mathbf{A}(\mathbf{x}^{k+1}-\mathbf{x}^k),\mathbf{y}-\mathbf{y}^{k+1}\rangle \\
& +\left(\frac{1}{2\tau_k}-\frac{L}{2}\right)\|\mathbf{x}^k-\mathbf{x}^{k+1}\|^2+\sup_{\mathbf{y}\in\mathcal{Y}'}\frac{1}{2\alpha_k}\|\mathbf{y}^{k+1}-\mathbf{y}\|^2+\left(\frac{1}{2\alpha_k}-\frac{\theta_k B}{2\mu_k}\right)\|\mathbf{y}^k-\mathbf{y}^{k+1}\|^2 \\
& +\sup_{\mathbf{x}\in\mathcal{X}'}\left(\frac{1}{2\tau_k}+\frac{\gamma}{2}\right)\|\mathbf{x}^{k+1}-\mathbf{x}\|^2+\sup_{\mathbf{x}\in\mathcal{X}'}-\langle \boldsymbol{\varepsilon}^k,\mathbf{x}-\widetilde{\mathbf{x}}^{k+1}\rangle-\tau_k\|\boldsymbol{\varepsilon}^k\|^2.
\end{aligned} \tag{S-34}
$$

We choose $\mu_k=\theta_k\alpha_k B$ and take expectation on the RHS of (S-34) w.r.t. $\boldsymbol{\xi}^k$ by conditioning on $\mathcal{F}_k$. Based on (S-13), we have

$$
\begin{aligned}
& -\inf_{\mathbf{y}\in\mathcal{Y}'}\theta_k\langle \mathbf{A}(\mathbf{x}^k-\mathbf{x}^{k-1}),\mathbf{y}-\mathbf{y}^k\rangle+\frac{\theta_k^2\alpha_k B^2}{2}\|\mathbf{x}^k-\mathbf{x}^{k-1}\|^2+\sup_{\mathbf{x}\in\mathcal{X}'}\frac{1}{2\tau_k}\|\mathbf{x}^k-\mathbf{x}\|^2 \\
& +\sup_{\mathbf{y}\in\mathcal{Y}'}\frac{1}{2\alpha_k}\|\mathbf{y}^k-\mathbf{y}\|^2\geq \mathbb{E}_{\boldsymbol{\xi}_k}\left[G_{\mathcal{X}',\mathcal{Y}'}(\mathbf{x}^{k+1},\mathbf{y}^{k+1})\,\middle|\,\mathcal{F}_k\right]-\tau_k\mathbb{E}_{\boldsymbol{\xi}^k}\left[\|\boldsymbol{\varepsilon}^k\|^2\,\middle|\,\mathcal{F}_k\right] \\
& +\frac{1}{\theta_{k+1}}\mathbb{E}_{\boldsymbol{\xi}_k}\Bigg[-\inf_{\mathbf{y}\in\mathcal{Y}'}\theta_{k+1}\langle \mathbf{A}(\mathbf{x}^{k+1}-\mathbf{x}^k),\mathbf{y}-\mathbf{y}^{k+1}\rangle+\theta_{k+1}\left(\frac{1}{2\tau_k}-\frac{L}{2}\right)\|\mathbf{x}^{k+1}-\mathbf{x}^k\|^2 \\
& +\sup_{\mathbf{x}\in\mathcal{X}'}\theta_{k+1}\left(\frac{1}{2\tau_k}+\frac{\gamma}{2}\right)\|\mathbf{x}^{k+1}-\mathbf{x}\|^2+\sup_{\mathbf{y}\in\mathcal{Y}'}\frac{\theta_{k+1}}{2\alpha_k}\|\mathbf{y}^{k+1}-\mathbf{y}\|^2\,\Bigg|\,\mathcal{F}_k\Bigg].
\end{aligned} \tag{S-35}
$$

Substitute (15), (16) and (17) into (S-35) and multiply both sides by α_k/α_0,

$$
\begin{aligned}
& \frac{\alpha_k}{\alpha_0}\Bigg[-\inf_{\mathbf{y}\in\mathcal{Y}'}\theta_k\langle \mathbf{A}(\mathbf{x}^k-\mathbf{x}^{k-1}),\mathbf{y}-\mathbf{y}^k\rangle+\frac{\theta_k^2\alpha_k B^2}{2}\|\mathbf{x}^k-\mathbf{x}^{k-1}\|^2+\sup_{\mathbf{x}\in\mathcal{X}'}\frac{1}{2\tau_k}\|\mathbf{x}^k-\mathbf{x}\|^2 \\
& +\sup_{\mathbf{y}\in\mathcal{Y}'}\frac{1}{2\alpha_k}\|\mathbf{y}^k-\mathbf{y}\|^2\Bigg]\geq\frac{\alpha_k}{\alpha_0}\,\mathbb{E}_{\boldsymbol{\xi}_k}\left[G_{\mathcal{X}',\mathcal{Y}'}(\mathbf{x}^{k+1},\mathbf{y}^{k+1})\,\middle|\,\mathcal{F}_k\right] \\
& +\frac{\alpha_{k+1}}{\alpha_0}\,\mathbb{E}_{\boldsymbol{\xi}_k}\Bigg[-\inf_{\mathbf{y}\in\mathcal{Y}'}\theta_{k+1}\langle \mathbf{A}(\mathbf{x}^{k+1}-\mathbf{x}^k),\mathbf{y}-\mathbf{y}^{k+1}\rangle+\sup_{\mathbf{x}\in\mathcal{X}'}\frac{\|\mathbf{x}^{k+1}-\mathbf{x}\|^2}{2\tau_{k+1}} \\
& +\frac{\theta_{k+1}^2\alpha_{k+1}B^2}{2}\|\mathbf{x}^{k+1}-\mathbf{x}^k\|^2+\sup_{\mathbf{y}\in\mathcal{Y}'}\frac{\|\mathbf{y}^{k+1}-\mathbf{y}\|^2}{2\alpha_{k+1}}\,\Bigg|\,\mathcal{F}_k\Bigg]-\frac{\alpha_k\tau_k}{\alpha_0}\,\mathbb{E}_{\boldsymbol{\xi}^k}\left[\|\boldsymbol{\varepsilon}^k\|^2\,\middle|\,\mathcal{F}_k\right].
\end{aligned} \tag{S-36}
$$

By Lemma S-1, we have

$$
\begin{aligned}
\frac{\alpha_k\tau_k}{\alpha_0} &\leq \left(1+\frac{\gamma k}{\alpha_0 B^2}\right)\frac{2B^2+2L+\gamma}{(2B^2+2L+\gamma)(B^2\alpha_0+L)+B^2\gamma k} \\
&\leq \frac{(\alpha_0 B^2+\gamma)(2B^2+2L+\gamma)}{\alpha_0\gamma B^4}=\bar{c}_1.
\end{aligned}
$$

Also, $\mathbf{z}^0 = \mathbf{x}^0$ implies $\mathbf{x}^0 = \mathbf{x}^{-1}$. Now we telescope (S-36) over $k = 0, 1, \ldots, K-1$ to obtain,

$$\sup_{\mathbf{x}\in\mathcal{X}'} \frac{1}{2\tau_0}\|\mathbf{x}^0 - \mathbf{x}\|^2 + \sup_{\mathbf{y}\in\mathcal{Y}'} \frac{1}{2\alpha_0}\|\mathbf{y}^0 - \mathbf{y}\|^2 \geq \mathbb{E}_{\Xi_K}\left[\sum_{k=0}^{K-1} \frac{\alpha_k}{\alpha_0} G_{\mathcal{X}',\mathcal{Y}'}(\mathbf{x}^{k+1}, \mathbf{y}^{k+1})\right]$$
$$+ \frac{\alpha_K}{\alpha_0}\, \mathbb{E}_{\Xi_K}\Big[- \inf_{\mathbf{y}\in\mathcal{Y}'} \theta_K \langle \mathbf{A}(\mathbf{x}^K - \mathbf{x}^{K-1}), \mathbf{y} - \mathbf{y}^K\rangle + \frac{\theta_K^2 \alpha_K B^2}{2}\|\mathbf{x}^K - \mathbf{x}^{K-1}\|^2$$
$$+ \sup_{\mathbf{y}\in\mathcal{Y}'} \frac{1}{2\alpha_K}\|\mathbf{y}^K - \mathbf{y}\|^2 + \sup_{\mathbf{x}\in\mathcal{X}'} \frac{1}{2\tau_K}\|\mathbf{x}^K - \mathbf{x}\|^2\Big] - \overline{c}_1 \sum_{k=0}^{K-1} \mathbb{E}_{\boldsymbol{\xi}_k}\left[\|\boldsymbol{\varepsilon}^k\|^2 \,|\, \mathcal{F}_k\right].$$

Using Young's inequality, we have

$$\inf_{\mathbf{y}\in\mathcal{Y}'} \langle \theta_K \mathbf{A}(\mathbf{x}^K - \mathbf{x}^{K-1}), \mathbf{y} - \mathbf{y}^K\rangle \leq \frac{\theta_K^2 \alpha_K B^2}{2}\|\mathbf{x}^K - \mathbf{x}^{K-1}\|^2 + \sup_{\mathbf{y}\in\mathcal{Y}'} \frac{1}{2\alpha_K}\|\mathbf{y} - \mathbf{y}^K\|.$$

Since we use option II in Algorithm 1, we have

$$S_K = \sum_{k=0}^{K-1} \frac{\alpha_k}{\alpha_0}, \quad \overline{\mathbf{x}}^K = \frac{1}{S_K}\sum_{k=1}^{K} \frac{\alpha_{k-1}}{\alpha_0}\mathbf{x}^k, \quad \overline{\mathbf{y}}^K = \frac{1}{S_K}\sum_{k=1}^{K} \frac{\alpha_{k-1}}{\alpha_0}\mathbf{y}^k.$$

From the joint convexity of $G_{\mathcal{X}',\mathcal{Y}'}(\cdot,\cdot)$, we employ Jensen's inequality to obtain

$$S_K \frac{1}{S_K}\sum_{k=0}^{K-1} \frac{\alpha_k}{\alpha_0} G_{\mathcal{X}',\mathcal{Y}'}(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}) \geq S_K G_{\mathcal{X}',\mathcal{Y}'}(\overline{\mathbf{x}}^{k+1}, \overline{\mathbf{y}}^{k+1}). \tag{S-37}$$

As a result,

$$S_K \mathbb{E}_{\Xi_K}\left[G_{\mathcal{X}',\mathcal{Y}'}(\overline{\mathbf{x}}^{k+1}, \overline{\mathbf{y}}^{k+1})\right] + \frac{\alpha_K}{2\alpha_0\tau_K}\mathbb{E}_{\Xi_K}\left[\sup_{\mathbf{x}\in\mathcal{X}'} \|\mathbf{x}^K - \mathbf{x}\|^2\right]$$
$$\leq \frac{1}{2\tau_0} R^2_{\mathcal{X}'}(\mathbf{x}^0) + \frac{1}{2\alpha_0} R^2_{\mathcal{Y}'}(\mathbf{y}^0) + \overline{c}_1 \sum_{k=0}^{K-1} \mathbb{E}_{\boldsymbol{\xi}_k}\left[\|\boldsymbol{\varepsilon}^k\|^2 \,|\, \mathcal{F}_k\right]. \tag{S-38}$$

From Lemma S-1, we have for any $K \geq 2$,

$$\frac{S_K}{K^2} \geq \frac{1}{K^2}\sum_{k=0}^{K-1} 1 + \frac{\gamma}{\alpha_0(2B^2 + 2L + \gamma)}k = \frac{1}{K^2}\left(K + \frac{\gamma}{2\alpha_0(2B^2 + 2L + \gamma)}(K-1)K\right)$$
$$\geq \frac{\gamma}{2\alpha_0(2B^2 + 2L + \gamma)}\frac{K-1}{K} \geq \frac{\gamma}{4\alpha_0(2B^2 + 2L + \gamma)}.$$

Thus, for any $K \geq 1$, $S_K \geq K^2/\max\left\{4\alpha_0(2B^2 + 2L + \gamma)/\gamma, 1\right\} = K^2/\overline{c}_1'$. Using $\mathbb{E}_{\boldsymbol{\xi}^k}\left[\|\boldsymbol{\varepsilon}^k\|^2 \,|\, \mathcal{F}_k\right] \leq \sigma^2$ in Assumption 1, we have

$$\mathbb{E}_{\Xi_K}\left[G_{\mathcal{X}',\mathcal{Y}'}(\overline{\mathbf{x}}^{k+1}, \overline{\mathbf{y}}^{k+1})\right] \leq \frac{\overline{c}_1'}{2K^2}\left(\frac{1}{\tau_0} R^2_{\mathcal{X}'}(\mathbf{x}^0) + \frac{1}{\alpha_0} R^2_{\mathcal{Y}'}(\mathbf{y}^0)\right) + \frac{\overline{c}_1 \overline{c}_1' \sigma^2}{K}. \tag{S-39}$$

Now (26) follows from (S-39) by defining $c_1' \triangleq \overline{c}_1'/2$ and $c_1 \triangleq \overline{c}_1 \overline{c}_1'$.

S-4.2 Proof of Part (ii)

Since $\mathbf{x}^*$ is the (unique) minimizer of (1), there exists $\mathbf{y}^* \in \mathbb{R}^m$ such that $(\mathbf{x}^*, \mathbf{y}^*)$ is a saddle point of (4). Choose $\mathcal{X}' = \{\mathbf{x}^*\}$ and $\mathcal{Y}' = \{\mathbf{y}^*\}$. By definition, $\mathbb{E}_{\Xi_K}\left[G_{\mathcal{X}',\mathcal{Y}'}(\overline{\mathbf{x}}^{k+1}, \overline{\mathbf{y}}^{k+1})\right] \geq 0$, for any $K \geq 1$. Thus (S-38) becomes

$$\frac{\alpha_K}{\tau_K}\mathbb{E}_{\Xi_K}\left[\|\mathbf{x}^K - \mathbf{x}^*\|^2\right] \leq \frac{\alpha_0}{\tau_0}\|\mathbf{x}^0 - \mathbf{x}^*\|^2 + \|\mathbf{y}^0 - \mathbf{y}^*\|^2 + 2\alpha_0 \overline{c}_1 \sum_{k=0}^{K-1} \mathbb{E}_{\boldsymbol{\xi}_k}\left[\|\boldsymbol{\varepsilon}^k\|^2 \,|\, \mathcal{F}_k\right].$$

From Lemma S-1, we have for any $K \geq 1$,

$$\begin{aligned}\frac{1}{K^2}\frac{\alpha_K}{\tau_K} &\geq \frac{1}{K^2}\left(\alpha_0 + \frac{\gamma}{2B^2+2L+\gamma}K\right)\left(B^2\alpha_0 + L + \frac{B^2\gamma}{2B^2+2L+\gamma}K\right)\\ &\geq \frac{B^2\gamma^2}{(2B^2+2L+\gamma)^2} = \frac{1}{c_2}.\end{aligned}$$

Based on $\mathbb{E}_{\boldsymbol{\xi}_k}\left[\|\boldsymbol{\varepsilon}^k\|^2 \,|\, \mathcal{F}_k\right] \leq \sigma^2$ in Assumption 1, we have for any $K \geq 1$, Combining these, we have

$$\mathbb{E}_{\Xi_K}\left[\|\mathbf{x}^K - \mathbf{x}^*\|^2\right] \leq \frac{c_2}{K^2}\left(\frac{\alpha_0}{\tau_0}\|\mathbf{x}^0 - \mathbf{x}^*\|^2 + \|\mathbf{y}^0 - \mathbf{y}^*\|^2\right) + \frac{2\alpha_0\bar{c}_1 c_2\sigma^2}{K}. \tag{S-40}$$

Then (28) follows from (S-40) by defining $c_2' \triangleq 2\bar{c}_1 c_2$.

S-5 Proof of Lemma S-1

It suffices to show (S-29), since (S-30) straightforwardly follow from (18). From (19), for any $k \geq 0$, we have

$$\begin{aligned}\alpha_{k+1} &= \frac{\sqrt{L^2 + 4B^2(B^2\alpha_k^2 + (L+\gamma)\alpha_k)} - L}{2B^2}\\ &= \frac{2\left(B^2\alpha_k^2 + (L+\gamma)\alpha_k\right)}{\sqrt{L^2 + 4B^2(B^2\alpha_k^2 + (L+\gamma)\alpha_k)} + L}.\end{aligned} \tag{S-41}$$

Since

$$\begin{aligned}2B^2\alpha_k + L = \sqrt{L^2 + 4B^2(B^2\alpha_k^2 + L\alpha_k)} &\leq \sqrt{L^2 + 4B^2(B^2\alpha_k^2 + (L+\gamma)\alpha_k)}\\ &\leq \sqrt{(L+\gamma)^2 + 4B^2(B^2\alpha_k^2 + (L+\gamma)\alpha_k)} = 2B^2\alpha_k + L + \gamma,\end{aligned}$$

from (S-41), we have

$$\alpha_k + \frac{\gamma\alpha_k}{2B^2\alpha_k + 2L + \gamma} \leq \alpha_{k+1} \leq \alpha_k + \frac{\gamma\alpha_k}{B^2\alpha_k + L} \leq \alpha_k + \frac{\gamma}{B^2}. \tag{S-42}$$

Since $\alpha_0 \geq 1$, then $\{\alpha_k\}_{k\geq 0}$ is strictly increasing and $\alpha_k \geq 1$, for any $k \geq 0$. As a result,

$$\alpha_k + \frac{\gamma\alpha_k}{2B^2\alpha_k + 2L + \gamma} \geq \alpha_k + \frac{\gamma}{2B^2 + 2L + \gamma}. \tag{S-43}$$

Combining (S-42) and (S-43), we obtain (S-29).

References

[1] M. Muresan, *A Concrete Approach to Classical Analysis*. Springer-Verlag New York, 2009.