E-companion: “Stochastic Compositional Optimization with Compositional Constraints”

EC.1 Proof of Results in Section 2

Letting Assumptions 1, 2, and 3 hold and letting $(x^*, \lambda, \pi_{1:2}, v) \in \mathcal{X} \times \mathbb{R}_+^m \times \Pi \times \mathcal{V}$ be a general feasible point, we first introduce two technical lemmas to facilitate our analysis.

Lemma EC.1 (Lemma 3.8 of Lan (2020)) *Assume function g is μ-strongly convex with respect to some Bergman distance V, i.e., $g(y) - g(\bar{y}) - g'(\bar{y})^\top (y - \bar{y}) \geq \mu V(y, \bar{y})$. If $\hat{y} \in \operatorname{argmin}_{y \in Y} \{\pi^\top y + g(y) + \tau V(y, \bar{y})\}$, then*

$$(\hat{y} - y)^\top \pi + g(\hat{y}) - g(y) \leq \tau V(y, \bar{y}) - (\tau + \mu) V(y, \hat{y}) - \tau V(\hat{y}, \bar{y}).$$

EC.1.1 Proof of Theorem 1

Proof: Recall that $\mathcal{L}(x, \pi_{1:2}, v, \lambda) = \mathcal{L}_F(x, \pi_{1:2}) + \lambda^\top \mathcal{L}_g(x, v)$ for any $(x, \lambda, \pi_{1:2}, v) \in \mathcal{X} \times \mathbb{R}_+^m \times \Pi \times \mathcal{V}$, we consider $Q(z_t, z)$ and decompose it as

$$\begin{aligned} Q(z_t, z) &= \mathcal{L}(x_t, \lambda, \pi_{1:2}, v) - \mathcal{L}(x^*, \lambda_t, \pi_{1:2,t}, v_t) \\ &= \mathcal{L}_F(x_t, \pi_{1:2}) - \mathcal{L}_F(x^*, \pi_{1:2,t}) + \lambda^\top \mathcal{L}_g(x_t, v) - \lambda_t^\top \mathcal{L}_g(x^*, v_t) \\ &= \mathcal{L}_F(x_t, \pi_{1:2}) - \mathcal{L}_F(x^*, \pi_{1:2,t}) + \lambda^\top \Big(\mathcal{L}_g(x_t, v) - \mathcal{L}_g(x_t, v_t) \Big) \\ &\quad + \lambda_t^\top \Big(\mathcal{L}_g(x_t, v_t) - \mathcal{L}_g(x^*, v_t) \Big) + (\lambda - \lambda_t)^\top \mathcal{L}_g(x_t, v_t) \\ &= \mathcal{L}_F(x_t, \pi_{1:2}) - \mathcal{L}_F(x_t, \pi_{1:2,t}) + \mathcal{L}_F(x_t, \pi_{1:2,t}) - \mathcal{L}_F(x^*, \pi_{1:2,t}) \\ &\quad + \lambda^\top \Big(\mathcal{L}_g(x_t, v) - \mathcal{L}_g(x_t, v_t) \Big) + \lambda_t^\top \Big(\mathcal{L}_g(x_t, v_t) - \mathcal{L}_g(x^*, v_t) \Big) + (\lambda - \lambda_t)^\top \mathcal{L}_g(x_t, v_t). \end{aligned} \tag{EC.1}$$

We then provide bounds for the terms above in expectation. First, by Lemma EC.3 in Appendix Section EC.1.2, we have

$$\begin{aligned} &\sum_{t=1}^{K} \mathbb{E}\Big[\mathcal{L}_F(x_t, \pi_{1:2}) - \mathcal{L}_F(x_t, \pi_{1:2,t}) \Big] \\ &\quad \leq \frac{40 K C_{f_1}^2 C_{f_2}^2}{\eta_0} + \sum_{t=1}^{K} \frac{\eta_{t-1}}{10} \mathbb{E}[\|x_t - x_{t-1}\|^2] \quad + \sqrt{K} L_{f_1} \sigma_{f_2}^2 + 3\sqrt{K} C_{f_1} \sigma_{f_2}. \end{aligned} \tag{EC.2}$$

Secondly, by Lemma EC.4 in Appendix Section EC.1.3, we obtain that

$$\begin{aligned} &\sum_{t=1}^{K} \mathbb{E}\Big[\mathcal{L}_F(x_t, \pi_{1:2,t}) - \mathcal{L}_F(x^*, \pi_{1:2,t}) + \lambda_t^\top \Big(\mathcal{L}_g(x_t, v_t) - \mathcal{L}_g(x^*, v_t) \Big) \Big] \\ &\quad \leq \frac{\eta}{2} \|x_0 - x^*\|^2 + \sum_{t=1}^{K} \Big(\frac{5 C_{f_1}^2 C_{f_2}^2}{2\eta} - \frac{3\eta}{10} \mathbb{E}[\|x_{t-1} - x_t\|^2] \Big) \\ &\quad\quad + \sum_{t=1}^{K} \Big(\frac{\alpha}{4} \mathbb{E}[\|\lambda_t - \lambda_{t-1}\|^2] + \frac{D_X^2 C_g^2}{\alpha} + \frac{5 \mathbb{E}[\|\lambda_{t-1}\|^2] C_g^2}{2\eta} \Big). \end{aligned} \tag{EC.3}$$

Thirdly, by Lemma EC.5 in Appendix Section EC.1.4 and the condition that $\|\lambda\| \le M_\lambda$, we obtain

$$\mathbb{E}\Big[\sum_{t=1}^{K} \lambda^\top \Big(\mathcal{L}_g(x_t, v) - \mathcal{L}_g(x_t, v_t)\Big)\Big] \le \sum_{t=1}^{K} \frac{10C_g^2 M_\lambda^2}{\eta} + \frac{\eta \mathbb{E}[\|x_t - x_{t-1}\|^2]}{10}. \tag{EC.4}$$

Finally, by Lemma EC.6 in Appendix Section EC.1.5, we obtain that

$$\begin{aligned}
&\mathbb{E}\Big[\sum_{t=1}^{K} (\lambda - \lambda_t)^\top \mathcal{L}_g(x_t, v_t)\Big] + \frac{\alpha}{2}\mathbb{E}[\|\lambda_K - \lambda\|^2] \\
&\le \sqrt{K} M_\lambda \sigma_g + \sum_{t=1}^{K} \Big(\frac{5C_g^2 \mathbb{E}[\|\lambda - \lambda_t\|^2]}{2\eta} + \frac{\eta}{10}\mathbb{E}[\|x_{t-1} - x_t\|^2] - \frac{\alpha}{4}\mathbb{E}[\|\lambda_{t-1} - \lambda_t\|^2]\Big) \qquad \text{(EC.5)} \\
&\quad + \frac{\alpha}{2}\mathbb{E}[\|\lambda_0 - \lambda\|^2] + \frac{K\sigma_g^2}{\alpha}.
\end{aligned}$$

Summing (EC.1) over $t = 1, 2, \cdots, K$, taking expectations, and combining (EC.2), (EC.3), (EC.4), and (EC.5), we conclude that

$$\begin{aligned}
&\sum_{t=1}^{K} \mathbb{E}\Big(\mathcal{L}(x_t, \lambda, \pi_{1:2}, v) - \mathcal{L}(x^*, \lambda_t, \pi_{1:2,t}, v_t)\Big) + \frac{\alpha}{2}\mathbb{E}[\|\lambda_K - \lambda\|^2] \\
&\le \frac{95KC_{f_1}^2 C_{f_2}^2}{2\eta} + \sqrt{K} L_{f_1} \sigma_{f_2}^2 + 3\sqrt{K} C_{f_1} \sigma_{f_2} + \frac{\eta}{2}\|x_0 - x^*\|^2 + \frac{K}{\alpha} D_X^2 C_g^2 + \sqrt{K} M_\lambda \sigma_g \\
&\quad + \frac{\alpha}{2}\mathbb{E}[\|\lambda_0 - \lambda\|^2] + \frac{K\sigma_g^2}{\alpha} + \sum_{t=1}^{K} \Big(\frac{5C_g^2 \mathbb{E}[\|\lambda_{t-1}\|^2]}{2\eta} + \frac{10C_g^2 M_\lambda^2}{\eta} + \frac{5C_g^2 \mathbb{E}[\|\lambda - \lambda_t\|^2]}{2\eta}\Big),
\end{aligned}$$

which completes the proof. ∎

Lemma EC.2 (Lemma 2 of Zhang and Lan (2024)) *Let $\mathcal{F}_t$ be a filtration and $\{\delta_t\}_{t=1}^N$ be a martingale noise sequence such that $\delta_j \in \mathcal{F}_{t+1}$ for $j = 1, 2, \cdots, t$, $\mathbb{E}[\delta_t | \mathcal{F}_t] = 0$, and $\mathbb{E}[\|\delta_t\|^2 | \mathcal{F}_t] \le \sigma^2$. For any random variable $\pi \in \Pi$ correlated with $\{\delta_t\}_{t=1}^N$, suppose it is bounded such that $\|\pi\| \le M_\pi$ uniformly, then*

$$\mathbb{E}[\sum_{t=1}^{N} \hat{\pi}^\top \delta_t] \le \sqrt{N} M_\pi \sigma.$$

EC.1.2 Lemma EC.3 and Its Proof

Lemma EC.3 *Suppose Assumptions 1, 2 and 3 hold, and Algorithm 1 generates $\{(x_t, \lambda_t, \pi_{1:2,t}, v_t)\}_{t=1}^N$ by setting $\tau_t = t/2$ and $\eta_t = \eta$ for $t = 0, 1, \cdots, N$. Then for any integer $K \le N$, we have*

$$\begin{aligned}
&\sum_{t=1}^{K} \mathbb{E}\Big[\mathcal{L}_F(x_t, \pi_{1:2}) - \mathcal{L}_F(x_t, \pi_{1:2,t})\Big] \\
&\quad \le \sum_{t=1}^{K} \frac{\eta_{t-1}}{10}\mathbb{E}[\|x_t - x_{t-1}\|^2] + \frac{40KC_{f_1}^2 C_{f_2}^2}{\eta} + 2\sqrt{K} L_{f_1} \sigma_{f_2}^2 + 3\sqrt{K} C_{f_1} \sigma_{f_2}.
\end{aligned}$$

Proof: This result can be derived by combining Propositions 5 and 6 of **?**. Here we provide the proof sketch. By decomposing $\mathcal{L}_F(x_t, \pi_{1:2}) - \mathcal{L}_F(x_t, \pi_{1:2,t})$, we have

$$\begin{aligned}
&\mathcal{L}_F(x_t, \pi_{1:2}) - \mathcal{L}_F(x_t, \pi_{1:2,t}) \\
&= \mathcal{L}_{f_1}(x_t, \pi_1, \pi_2) - \mathcal{L}_{f_1}(x_t, \pi_1, \pi_{2,t}) + \mathcal{L}_{f_1}(x_t, \pi_1, \pi_{2,t}) - \mathcal{L}_{f_1}(x_t, \pi_{1,t}, \pi_{2,t}).
\end{aligned}$$

We provide bound for each term after the above decomposition. First, we consider $\mathcal{L}_{f_1}(x_t, \pi_1, \pi_2) - \mathcal{L}_{f_1}(x_t, \pi_1, \pi_{2,t})$ and obtain

$$\begin{aligned}
&\sum_{t=1}^{K} \mathbb{E}\Big(\mathcal{L}_{f_1}(x_t, \pi_1, \pi_2) - \mathcal{L}_{f_1}(x_t, \pi_1, \pi_{2,t})\Big) = \sum_{t=1}^{K} \mathbb{E}[\Big\langle \mathcal{L}_{f_2}(x_t, \pi_2) - \mathcal{L}_{f_2}(x_t, \pi_{2,t}), \pi_1 \Big\rangle] \\
&= \sum_{t=1}^{K} \mathbb{E}\Big[\Big\langle (\pi_2 - \pi_{2,t})^\top x_t - f_2^*(\pi_2) + f_2^*(\pi_{2,t}), \pi_1 \Big\rangle \Big] \\
&= \sum_{t=1}^{K} \mathbb{E}\Big[\Big\langle (\pi_2 - \pi_{2,t})^\top x_{t-1} - f_2^*(\pi_2) + f_2^*(\pi_{2,t}), \pi_1 \Big\rangle \Big] \\
&\quad + \sum_{t=1}^{K} \mathbb{E}\Big[\Big\langle (\pi_2 - \pi_{2,t})^\top (x_t - x_{t-1}), \pi_1 \Big\rangle \Big],
\end{aligned}$$

By recalling (13) that $\pi_{2,t} \in \operatorname{argmax}_{\pi_2 \in \Pi_2} \{\langle \pi_2, x_{t-1} \rangle - f_2^*(\pi_2)\}$ and using Assumptions 3 that $\pi_1 \geq 0$ for non-affine f_2, we obtain for both affine (π_2 is a constant) and non-affine f_2,

$$\mathbb{E}\Big[\Big\langle (\pi_2 - \pi_{2,t})^\top x_{t-1} - f_2^*(\pi_2) + f_2^*(\pi_{2,t}), \pi_1 \Big\rangle \Big] \leq 0,$$

which further implies that

$$\begin{aligned}
&\sum_{t=1}^{K} \mathbb{E}\Big(\mathcal{L}_{f_1}(x_t, \pi_1, \pi_2) - \mathcal{L}_{f_1}(x_t, \pi_1, \pi_{2,t})\Big) \leq \sum_{t=1}^{K} \mathbb{E}[\langle (\pi_2 - \pi_{2,t})\pi_1, x_t - x_{t-1} \rangle] \\
&\leq \sum_{t=1}^{K} \Big[\frac{5\mathbb{E}[\|(\pi_2 - \pi_{2,t})\pi_1\|^2]}{\eta} + \frac{\eta}{20}\mathbb{E}[\|x_t - x_{t-1}\|^2] \Big] \qquad \text{(EC.6)} \\
&\leq \sum_{t=1}^{K} \Big[\frac{20 C_{f_1}^2 C_{f_2}^2}{\eta} + \frac{\eta}{20}\mathbb{E}[\|x_t - x_{t-1}\|^2] \Big].
\end{aligned}$$

The next part is similar to Proposition 6 of **?**. Consider the following decomposition:

$$\begin{aligned}
&\mathbb{E}[\mathcal{L}_{f_1}(x_t, \pi_1, \pi_{2,t}) - \mathcal{L}_{f_1}(x_t, \pi_{1,t}, \pi_{2,t})] = \mathbb{E}[\langle \pi_1 - \pi_{1,t}, f_2(x_t) \rangle - \{f_1^*(\pi_1) - f_1^*(\pi_{1,t})\}] \\
&= \mathbb{E}[\Big\langle \pi_1 - \pi_{1,t}, f_2(x_{t-1}, \xi_{2,t}^1) \Big\rangle - \{f_1^*(\pi_1) - f_1^*(\pi_{1,t})\}] \\
&\quad + \mathbb{E}[\Big\langle \pi_{1,t} - \pi_{1,t-1}, f_2(x_{t-1}) - f_2(x_{t-1}, \xi_{2,t}^1) \Big\rangle] + \mathbb{E}[\langle \pi_1 - \pi_{1,t}, f_2(x_t) - f_2(x_{t-1}) \rangle] \qquad \text{(EC.7)} \\
&\quad + \mathbb{E}[\Big\langle \pi_1, f_2(x_{t-1}) - f_2(x_{t-1}, \xi_{2,t}^t) \Big\rangle] + \underbrace{\mathbb{E}[\Big\langle \pi_{1,t-1}, f_2(x_{t-1}) - f_2(x_{t-1}, \xi_{2,t}^1) \Big\rangle]}_{=0}.
\end{aligned}$$

Among these terms, the update rule of $\pi_{1,t}$, the strong convexity of f_1^* with respect to $D_{f_1^*}$, and the choice of τ_t imply

$$\begin{aligned}
&\mathbb{E}[\sum_{t=1}^{K}\left\langle \pi_1 - \pi_{1,t}, f_2(x_{t-1}, \xi_{2,t}^1)\right\rangle - \{f_1^*(\pi_1) - f_1^*(\pi_{1,t})\}] \\
&\quad \leq \mathbb{E}[\sum_{t=1}^{K} \tau_t D_{f_1^*}(\pi_1; \pi_{1,t-1}) - (\tau_t + 1) D_{f_1^*}(\pi_1; \pi_{1,t-1}) - \tau_t D_{f_1^*}(\pi_{1,t}; \pi_{1,t-1})\}] \\
&\quad \leq -\mathbb{E}[\sum_{t=1}^{K} \tau_t D_{f_1^*}(\pi_{1,t}; \pi_{1,t-1})\}].
\end{aligned}$$

The $1/L_{f_1}$-strong convexity of $D_{f_1^*}$, the choice of τ_t and the Cauchy-Schwartz inequality imply

$$\begin{aligned}
&\mathbb{E}\sum_{t=1}^{K}\Big[(\pi_{1,t} - \pi_{1,t-1})(f_2(x_{t-1}, \xi_{2,t}^1) - f_2(x_{t-1})) - \tau_t \mathcal{D}_{f_1^*}(\pi_{1,t}; \pi_{1,t-1})\Big] \\
&\quad \leq \sum_{t=2}^{K} \frac{L_{f_1}\sigma_{f_2}^2}{\tau_t} + \sqrt{\mathbb{E}[\|\pi_{1,1} - \pi_{1,0}\|^2]}\sqrt{\mathbb{E}[\Big\| f_2(x_0, \xi_{2,0}^1) - f_2(x_0)\Big\|^2]} \\
&\quad \leq 2L_{f_1}\sigma_{f_2}^2 \log K + 2C_{f_1}\sigma_{f_2} \\
&\quad \leq 2\sqrt{K} L_{f_1}\sigma_{f_2}^2 + 2\sqrt{K} C_{f_1}\sigma_{f_2}.
\end{aligned}$$

Lemma 2 of **?** implies that

$$\mathbb{E}\left[\sum_{t=1}^{K}\left\langle \pi_1, f_2(x_{t-1}) - f_2(x_{t-1}, \xi_{2,t}^t)\right\rangle\right] \leq \sqrt{K} C_{f_1}\sigma_{f_2}.$$

Similar to (EC.6), we also have

$$\mathbb{E}[\sum_{t=1}^{K}\langle \pi_1 - \pi_{1,t}, f_2(x_t) - f_2(x_{t-1})\rangle] \leq \sum_{t=1}^{K}\Big[\frac{20C_{f_1}^2 C_{f_2}^2}{\eta} + \frac{\eta}{20}\mathbb{E}[\|x_t - x_{t-1}\|^2]\Big].$$

Combining them with the decomposition in (EC.7), we obtain

$$\begin{aligned}
&\sum_{t=1}^{K}\mathbb{E}\Big(\mathcal{L}_{f_1}(x_t, \pi_1, \pi_{2,t}) - \mathcal{L}_{f_1}(x_t, \pi_{1,t}, \pi_{2,t})\Big) \\
&\quad \leq \sum_{t=1}^{K}\Big[\frac{20C_{f_1}^2 C_{f_2}^2}{\eta} + \frac{\eta}{20}\mathbb{E}[\|x_t - x_{t-1}\|^2]\Big] + 2\sqrt{K} L_{f_1}\sigma_{f_2}^2 + 3\sqrt{K} C_{f_1}\sigma_{f_2}.
\end{aligned}$$

The desired result then follows from adding to preceding inequality to (EC.6). ∎

EC.1.3 Lemma EC.4 and Its Proof

Lemma EC.4 *Suppose Assumptions 1, 2, and 3 hold, and Algorithm 1 generates* $\{(x_t, \lambda_t, \pi_{1:2,t}, v_t)\}_{t=1}^N$ *by setting* $\tau_t = t/2$*,* $\alpha_t = \alpha$*, and* $\eta_t = \eta$ *for all* $0 \le t \le N$*. We have*

$$\begin{aligned}
&\sum_{t=1}^K \mathbb{E}\Big[\mathcal{L}_F(x_t, \pi_{1:2,t}) - \mathcal{L}_F(x^*, \pi_{1:2,t}) + \lambda_t^\top\Big(\mathcal{L}_g(x_t, v_t) - \mathcal{L}_g(x^*, v_t)\Big)\Big] \\
&\quad \le \frac{\eta}{2}\mathbb{E}[\|x_{t-1} - x^*\|^2] - \sum_{t=1}^K \frac{3\eta}{10}\mathbb{E}\Big[\|x_{t-1} - x_t\|^2\Big] + \sum_{t=1}^K \frac{5C_{f_1}^2 C_{f_2}^2}{2\eta} \\
&\qquad + \sum_{t=1}^K \frac{\alpha}{4}\mathbb{E}[\|\lambda_t - \lambda_{t-1}\|^2] + \sum_{t=1}^K \frac{1}{\alpha} D_X^2 C_g^2 + \sum_{t=1}^K \frac{5\mathbb{E}[\|\lambda_{t-1}\|^2] C_g^2}{2\eta}].
\end{aligned}$$

Proof. Recall Algorithm 1, for $t = 1, 2, \cdots, K$, we note that $\pi_{2,t}\pi_{1,t}, v_t\lambda_t \in \mathbb{R}^{d_x}$ and obtain

$$\begin{aligned}
&\mathcal{L}_F(x_t, \pi_{1:2,t}) - \mathcal{L}_F(x^*, \pi_{1:2,t}) + \lambda_t^\top\Big(\mathcal{L}_g(x_t, v_t) - \mathcal{L}_g(x^*, v_t)\Big) \\
&\quad = \langle \pi_{2,t}\pi_{1,t}, x_t - x^* \rangle + \langle v_t\lambda_t, x_t - x^* \rangle \\
&\quad = \Big\langle \pi_{2,t}^0\pi_{1,t}^0, x_t - x^* \Big\rangle + \Big\langle v_t^0\lambda_{t-1}, x_t - x^* \Big\rangle \\
&\qquad + \Big\langle \pi_{2,t}\pi_{1,t} - \pi_{2,t}^0\pi_{1,t}^0, x_t - x^* \Big\rangle + \Big\langle v_t\lambda_t - v_t^0\lambda_{t-1}, x_t - x^* \Big\rangle,
\end{aligned}$$

and

$$x_t = \operatorname*{argmin}_{x \in \mathcal{X}} \Big\{ \Big\langle \pi_{2,t}^0\pi_{1,t}^0 + v_t^0\lambda_{t-1}, x \Big\rangle + \frac{\eta_{t-1}}{2}\|x_{t-1} - x\|^2 \Big\}.$$

By applying Lemma EC.1 to the above update rule, since $x^* \in \mathcal{X}$, we have

$$\Big\langle \pi_{2,t}^0\pi_{1,t}^0 + v_t^0\lambda_{t-1}, x_t - x^* \Big\rangle \le \frac{\eta}{2}\|x_{t-1} - x^*\|^2 - \frac{\eta}{2}\|x_{t-1} - x_t\|^2 - \frac{\eta}{2}\|x_t - x^*\|^2,$$

which further yields

$$\begin{aligned}
&\mathcal{L}_F(x_t, \pi_{1:2,t}) - \mathcal{L}_F(x^*, \pi_{1:2,t}) + \lambda_t^\top\Big(\mathcal{L}_g(x_t, v_t) - \mathcal{L}_g(x^*, v_t)\Big) \\
&\quad \le \frac{\eta}{2}\|x_{t-1} - x^*\|^2 - \frac{\eta}{2}\|x_{t-1} - x_t\|^2 - \frac{\eta}{2}\|x_t - x^*\| \\
&\qquad + \underbrace{\Big\langle \pi_{2,t}\pi_{1,t} - \pi_{2,t}^0\pi_{1,t}^0, x_t - x^* \Big\rangle}_{\Delta_0^t} + \underbrace{\Big\langle v_t\lambda_t - v_t^0\lambda_{t-1}, x_t - x^* \Big\rangle}_{\Lambda_0^t}.
\end{aligned} \tag{EC.8}$$

Then, we decompose Δ_0^t by

$$\Delta_0^t = \Big(\pi_{2,t}\pi_{1,t} - \pi_{2,t}^0\pi_{1,t}^0\Big)(x_{t-1} - x^*) + \Big(\pi_{2,t}\pi_{1,t} - \pi_{2,t}^0\pi_{1,t}^0\Big)(x_t - x_{t-1}).$$

We note that given x_{t-1}, y_t,

$$\mathbb{E}\Big[\pi_{2,t}\pi_{1,t} - \pi_{2,t}^0\pi_{1,t}^0 \Big| x_{t-1}, y_t\Big] = 0,$$

and

$$\begin{aligned}
&\mathbb{E}\Big[\|\pi_{2,t}\pi_{1,t} - \pi_{2,t}^0\pi_{1,t}^0\|^2 | x_{t-1}, y_t\Big] \\
&\quad \le \mathbb{E}\Big[\|\pi_{1,t} - \pi_{1,t}^0\|^2 \Big| x_{t-1}, y_t\Big]\mathbb{E}\Big[\|\pi_{2,t}^0\|^2 \Big| x_{t-1}\Big] + \mathbb{E}\Big[\|(\pi_{2,t}^0 - \pi_{2,t})\pi_{1,t}\|^2 | x_{t-1}, y_t\Big] \\
&\quad \le 2C_{f_1}^2 C_{f_2}^2.
\end{aligned}$$

By the independence between $(x_{t-1}-x^*)$ and the mean-zero term $(\pi_{2,t}\pi_{1,t}-\pi^0_{2,t}\pi^0_{1,t})$, we have $\mathbb{E}\Big[\Big(\pi_{2,t}\pi_{1,t}-\pi^0_{2,t}\pi^0_{1,t}\Big)^\top(x_{t-1}-x^*)\Big]=0$. Moreover, to handle the possible correlation between (x_t-x_{t-1}) and $(\pi_{2,t}\pi_{1,t}-\pi^0_{2,t}\pi^0_{1,t})$, we utilize the fact that $ab\le\frac{5a^2}{2}+\frac{b^2}{10}$ for all $a,b\in\mathbb{R}$, and obtain

$$\begin{aligned}\mathbb{E}\left[\sum_{t=1}^K\Delta_0^t\right]&=\mathbb{E}\Big[\sum_{t=1}^K\Big\langle\pi_{2,t}\pi_{1,t}-\pi^0_{2,t}\pi^0_{1,t},x_t-x_{t-1}\Big\rangle\Big]\\&\le\sum_{t=1}^K\Big(\frac{5C_{f_1}^2C_{f_2}^2}{2\eta}+\frac{\eta\mathbb{E}[\|x_t-x_{t-1}\|^2]}{10}\Big).\end{aligned}\tag{EC.9}$$

We then decompose Λ_0^t as

$$\Lambda_0^t=\underbrace{\langle v_t(\lambda_t-\lambda_{t-1}),x_t-x^*\rangle}_{\Lambda_1^t}+\underbrace{\Big\langle(v_t-v_t^0)\lambda_{t-1},x_t-x^*\Big\rangle}_{\Lambda_2^t}.$$

By the fact that $2ab\le a^2+b^2$ and Assumptions 1 and 2 that $\|x_t-x^*\|\le D_X$ and $\|v_t\|\le C_g$, for any $\alpha>0$, we have

$$\begin{aligned}\mathbb{E}[\Lambda_1^t]&\le\frac{\alpha}{4}\mathbb{E}[\|\lambda_t-\lambda_{t-1}\|^2]+\frac{1}{\alpha}\mathbb{E}[\|v_t^\top(x_t-x^*)\|^2]\\&\le\frac{\alpha}{4}\mathbb{E}[\|\lambda_t-\lambda_{t-1}\|^2]+\frac{1}{\alpha}D_X^2C_g^2.\end{aligned}\tag{EC.10}$$

Meanwhile, since $\mathbb{E}[v_t-v_t^0]=0$ and $\mathbb{E}[\|v_t-v_t^0\|^2]\le C_g^2$, we obtain

$$\begin{aligned}\mathbb{E}\left[\sum_{t=1}^K\Lambda_2^t\right]&=\mathbb{E}\Big[\sum_{t=1}^K\Big\langle(v_t-v_t^0)\lambda_{t-1},x_{t-1}-x^*\Big\rangle\Big]+\mathbb{E}\Big[\sum_{t=1}^K\Big\langle(v_t-v_t^0)\lambda_{t-1},x_t-x_{t-1}\Big\rangle\Big]\\&=\sum_{t=1}^K\mathbb{E}\Big[\Big\langle(v_t-v_t^0)\lambda_{t-1},x_t-x_{t-1}\Big\rangle\Big]\\&\le\sum_{t=1}^K\mathbb{E}\Big[\frac{5\|(v_t-v_t^0)\lambda_{t-1}\|^2}{2\eta}\Big]+\sum_{t=1}^K\frac{\eta\mathbb{E}[\|x_t-x_{t-1}\|^2]}{10}\\&\le\sum_{t=1}^K\frac{5C_g^2\mathbb{E}[\|\lambda_{t-1}\|^2]}{2\eta}]+\sum_{t=1}^K\frac{\eta\mathbb{E}[\|x_t-x_{t-1}\|^2]}{10},\end{aligned}\tag{EC.11}$$

where the second equality holds by the independence between $(v_t-v_t^0)\lambda_{t-1}$ and $(x_{t-1}-x^*)$, and the second last inequality holds due to the fact that $ab\le\frac{5a^2}{2}+\frac{b^2}{10}$. Taking expectations on both sides of (EC.8), summing over $t=1,\cdots,N$, and substituting (EC.9), (EC.10), and (EC.11) into it, we obtain

$$\begin{aligned}&\sum_{t=1}^K\mathbb{E}\Big[\mathcal{L}_F(x_t,\pi_{1:2,t})-\mathcal{L}_F(x^*,\pi_{1:2,t})+\lambda_t^\top\Big(\mathcal{L}_g(x_t,v_t)-\mathcal{L}_g(x^*,v_t)\Big)\Big]\\&\le\sum_{t=1}^K\mathbb{E}\Big[\frac{\eta}{2}\|x_t-x^*\|^2-\frac{\eta}{2}\|x_t-x_{t+1}\|^2-\frac{\eta}{2}\|x_{t+1}-x^*\|+\Delta_0^t+\Lambda_1^t+\Lambda_2^t\Big]\\&\le\frac{\eta}{2}\mathbb{E}[\|x_{t-1}-x^*\|^2]-\sum_{t=1}^K\frac{3\eta}{10}\mathbb{E}\Big[\|x_{t-1}-x_t\|^2\Big]+\sum_{t=1}^K\frac{5C_{f_1}^2C_{f_2}^2}{2\eta_{t-1}}\\&\quad+\sum_{t=1}^K\frac{\alpha}{4}\mathbb{E}[\|\lambda_t-\lambda_{t-1}\|^2]+\sum_{t=1}^K\frac{D_X^2C_g^2}{\alpha}+\sum_{t=1}^K\frac{5\mathbb{E}[\|\lambda_{t-1}\|^2]C_g^2}{2\eta}],\end{aligned}$$

where the last inequality holds by the fact that $\eta_t = \eta$ for $t = 1, 2, \cdots, N$. This completes the proof. ■

EC.1.4 Lemma EC.5 and Its Proof

Lemma EC.5 *Suppose Assumptions 1, 2 and 3 hold, and Algorithm 1 generates* $\{(x_t, \lambda_t, \pi_t, v_t)\}_{t=1}^N$ *by setting* $\tau_t = t/2$, $\alpha_t = \alpha$, *and* $\eta_t = \eta$ *for all* $0 \le t \le N$. *Let* $\lambda \in \mathbb{R}_+^m$ *be a nonnegative random variable whose norm is bounded such that* $\|\lambda\| \le M_\lambda$ *uniformly. Then for any* $K \le N$, *we have*

$$\mathbb{E}\Big[\sum_{t=1}^K \lambda^\top\Big(\mathcal{L}_g(x_t, v) - \mathcal{L}_g(x_t, v_t)\Big)\Big] \le \frac{10 M_\lambda^2 C_g^2}{\eta} + \frac{\eta \mathbb{E}[\|x_t - x_{t-1}\|^2]}{10}.$$

Proof. Recall that the update rule is that

$$v_t \in \operatorname*{argmax}_{v \in \mathcal{V}} \Big\{ \langle v, x_{t-1} \rangle - g^*(v) \Big\}.$$

By the fact that $\lambda \ge 0$, we have

$$\begin{aligned}
&\lambda^\top\Big(\mathcal{L}_g(x_t, v) - \mathcal{L}_g(x_t, v_t)\Big) \\
&= \lambda^\top\Big((v - v_t)^\top x_{t-1} - (g^*(v) - g^*(v_t)) + (v - v_t)^\top (x_t - x_{t-1})\Big) \\
&\le \lambda^\top\Big((v - v_t)^\top (x_t - x_{t-1})\Big).
\end{aligned}$$

By taking expectations on both sides of the above inequality, we have

$$\begin{aligned}
\mathbb{E}\left[\lambda^\top\Big(\mathcal{L}_g(x_t, v) - \mathcal{L}_g(x_t, v_t)\Big)\right] &\le \mathbb{E}\Big[\frac{5\|(v - v_t)\lambda\|^2}{2\eta} + \frac{\eta\|x_t - x_{t-1}\|^2}{10}\Big] \\
&\le \frac{5M_\lambda^2 \mathbb{E}[\|v - v_t\|^2]}{2\eta_{t-1}} + \frac{\eta \mathbb{E}[\|x_t - x_{t-1}\|^2]}{10} \le \frac{10 C_g^2 M_\lambda^2}{\eta} + \frac{\eta \mathbb{E}[\|x_t - x_{t-1}\|^2]}{10},
\end{aligned}$$

where the first inequality uses the fact that $ab \le \frac{5a^2}{2} + \frac{b^2}{10}$, the second inequality holds by the condition that $\|\lambda\| \le M_\lambda$, and the third inequality holds by Assumption 2 that $\|v - v_t\|^2 \le 2\|v\|^2 + 2\|v_t\|^2 \le 4C_g^2$. Summing the above inequality over $t = 1, \cdots, K$, we obtain

$$\mathbb{E}\Big[\sum_{t=1}^K \lambda^\top\Big(\mathcal{L}_g(x_t, v) - \mathcal{L}_g(x_t, v_t)\Big)\Big] \le \frac{10 C_g^2 M_\lambda^2}{\eta} + \frac{\eta \mathbb{E}[\|x_t - x_{t-1}\|^2]}{10},$$

which completes the proof. ■

EC.1.5 Lemma EC.6 and Its Proof

Lemma EC.6 *Suppose Assumptions 1, 2, and 3 hold, and Algorithm 1 generates* $\{(x_t, \lambda_t, \pi_{1:2,t}, v_t)\}_{t=1}^N$ *by setting* $\tau_t = t/2$, $\alpha_t = \alpha$, *and* $\eta_t = \eta$ *for all* $0 \le t \le N$. *Let* $\lambda \in \mathbb{R}_+^m$ *be a bounded nonnegative* *random variable such that $\|\lambda\| \le M_\lambda$. For any integer $K \le N$, we have*

$$\mathbb{E}\Big[\sum_{t=1}^{K}(\lambda-\lambda_t)^\top \mathcal{L}_g(x_t,v_t)\Big] + \frac{\alpha_{K-1}}{2}\mathbb{E}[\|\lambda_K-\lambda\|^2]$$
$$\le \sqrt{K}M_\lambda\sigma_g + \sum_{t=1}^{K}\Big(\frac{5C_g^2\mathbb{E}[\|\lambda-\lambda_t\|^2]}{2\eta} + \frac{\eta}{10}\mathbb{E}[\|x_{t-1}-x_t\|^2] - \frac{\alpha}{4}\mathbb{E}[\|\lambda_{t-1}-\lambda_t\|^2]\Big)$$
$$+ \frac{\alpha}{2}\mathbb{E}[\|\lambda_0-\lambda\|^2] + \frac{K\sigma_g^2}{\alpha}.$$

Proof: By decomposing $(\lambda-\lambda_t)^\top\mathcal{L}_g(x_t,v_t)$, we have,

$$\begin{aligned}
&(\lambda-\lambda_t)^\top\mathcal{L}_g(x_t,v_t)\\
&\quad= (\lambda-\lambda_t)^\top\Big(\mathcal{L}_g(x_t,v_t)-\mathcal{L}_g(x_{t-1},v_t)\Big) + (\lambda-\lambda_t)^\top[\mathcal{L}_g(x_{t-1},v_t)-g(x_{t-1},\zeta_{t-1}^1)]\\
&\qquad + (\lambda-\lambda_t)g(x_{t-1},\zeta_{t-1}^1)\\
&= \underbrace{\langle v_t(\lambda-\lambda_t), x_{t-1}-x_t\rangle}_{T_{1,t}} + \underbrace{(\lambda-\lambda_t)^\top[g(x_{t-1})-g(x_{t-1},\zeta_{t-1}^1)]}_{T_{2,t}}\\
&\qquad + \underbrace{(\lambda-\lambda_t)^\top g(x_{t-1},\zeta_{t-1}^1)}_{T_{3,t}}.
\end{aligned} \tag{EC.12}$$

We provide bounds for $T_{1,t}, T_{2,t}$, and $T_{3,t}$ in our analysis.

First, for $T_{1,t}$, by using the fact that $\langle a,b\rangle \le \frac{5\|a\|^2}{2} + \frac{\|b\|^2}{10}$ and Assumption 2 that $\|v_t\|^2 \le C_g^2$, we obtain

$$\begin{aligned}
\mathbb{E}[\langle v_t(\lambda-\lambda_t), x_{t-1}-x_t\rangle] &\le \frac{5\mathbb{E}[\|v_t(\lambda-\lambda_t)\|^2]}{2\eta} + \frac{\eta\mathbb{E}[\|x_{t-1}-x_t\|^2]}{10}\\
&\le \frac{5C_g^2\mathbb{E}[\|\lambda-\lambda_t\|^2]}{2\eta} + \frac{\eta\mathbb{E}[\|x_{t-1}-x_t\|^2]}{10}.
\end{aligned} \tag{EC.13}$$

Second, consider $T_{2,t}$, by the independence between λ_{t-1} and the fact that $\mathbb{E}[g(x_{t-1})-g(x_{t-1},\zeta_{t-1}^1)] = 0$, we have

$$\mathbb{E}\Big[\lambda_{t-1}^\top(g(x_{t-1})-g(x_{t-1},\zeta_{t-1}^1))\Big] = 0,$$

which further implies that

$$\begin{aligned}
\mathbb{E}[T_{2,t}] &= \mathbb{E}\Big[(\lambda-\lambda_{t-1})^\top[g(x_{t-1})-g(x_{t-1},\zeta_{t-1}^1)]\Big]\\
&\quad + \mathbb{E}\Big[(\lambda_{t-1}-\lambda_t)^\top[g(x_{t-1})-g(x_{t-1},\zeta_{t-1}^1)]\Big]\\
&= \mathbb{E}\Big[\lambda^\top[g(x_{t-1})-g(x_{t-1},\zeta_{t-1}^1)]\Big] + \mathbb{E}\Big[(\lambda_{t-1}-\lambda_t)^\top[g(x_{t-1})-g(x_{t-1},\zeta_{t-1}^1)]\Big].
\end{aligned} \tag{EC.14}$$

Recall that λ is random but bounded such that $\|\lambda\| \le M_\lambda$, by Lemma EC.2 and Assumption 2 that $\mathbb{E}[\|g(x_{t-1})-g(x_{t-1},\zeta_{t-1}^1)\|^2] \le \sigma_g^2$, we obtain

$$\mathbb{E}\Big[\sum_{t=1}^{K}\lambda^\top\Big(g(x_{t-1})-g(x_{t-1},\zeta_{t-1}^1)\Big)\Big] \le \sqrt{K}M_\lambda\sigma_g. \tag{EC.15}$$

Meanwhile, by the fact that $\langle a, b\rangle \le \frac{\alpha\|a\|^2}{4} + \frac{\|b\|^2}{\alpha}$ for $\alpha > 0$, we bound the second term within the right side of (EC.14) by

$$\begin{aligned}&\mathbb{E}\Big[(\lambda_{t-1}-\lambda_t)^\top[g(x_{t-1})-g(x_{t-1},\zeta_{t-1}^1)]\Big]\\&\le \frac{\alpha\mathbb{E}[\|\lambda_{t-1}-\lambda_t\|^2]}{4}+\frac{\mathbb{E}[\|g(x_{t-1})-g(x_{t-1},\zeta_{t-1}^1)\|^2]}{\alpha}\le \frac{\alpha\mathbb{E}[\|\lambda_{t-1}-\lambda_t\|^2]}{4}+\frac{\sigma_g^2}{\alpha}.\end{aligned} \tag{EC.16}$$

Summing (EC.14) over $t = 1, \cdots, K$ and applying (EC.15) and (EC.16), we obtain

$$\sum_{t=1}^{K}\mathbb{E}[T_{2,t}] \le \sqrt{K}M_\lambda\sigma_g + \sum_{t=1}^{K}\Big(\frac{\alpha\mathbb{E}[\|\lambda_{t-1}-\lambda_t\|^2]}{4}+\frac{\sigma_g^2}{\alpha}\Big). \tag{EC.17}$$

Third, consider $T_{3,t}$, by recalling (15) that

$$\lambda_t \in \operatorname*{argmax}_{\lambda\in\mathbb{R}_+^m}\Big\{\Big\langle g(x_{t-1},\zeta_{t-1}^1),\lambda\Big\rangle - \frac{\alpha}{2}\|\lambda_t-\lambda\|^2\Big\},$$

and by Lemma EC.1, we have

$$\begin{aligned}T_{3,t} &= (\lambda-\lambda_t)^\top g(x_{t-1},\zeta_{t-1}^1) = -(\lambda_t-\lambda)^\top g(x_{t-1},\zeta_{t-1}^1)\\&\le \frac{\alpha}{2}\|\lambda_{t-1}-\lambda\|^2 - \frac{\alpha}{2}\|\lambda_{t-1}-\lambda_t\|^2 - \frac{\alpha}{2}\|\lambda_t-\lambda\|^2.\end{aligned} \tag{EC.18}$$

Finally, summing (EC.12) over $t = 1, \cdots, N$, taking expectations on both sides, and plugging in (EC.13), (EC.17), and (EC.18), we conclude that

$$\begin{aligned}&\mathbb{E}\Big[\sum_{t=1}^{K}(\lambda-\lambda_t)^\top\mathcal{L}_g(x_t,v_t)\Big] + \frac{\alpha}{2}\mathbb{E}[\|\lambda_K-\lambda\|^2]\\&\le \sqrt{K}M_\lambda\sigma_g + \sum_{t=1}^{K}\Big(\frac{5C_g^2\mathbb{E}[\|\lambda-\lambda_t\|^2]}{2\eta}+\frac{\eta}{10}\mathbb{E}[\|x_{t-1}-x_t\|^2]-\frac{\alpha}{4}\mathbb{E}[\|\lambda_{t-1}-\lambda_t\|^2]\Big)\\&\quad+\frac{\alpha}{2}\mathbb{E}[\|\lambda_0-\lambda\|^2]+\frac{K\sigma_g^2}{\alpha},\end{aligned}$$

which completes the proof. ∎

EC.1.6 Proof of Proposition 1

By applying Theorem 1 with $\lambda = \lambda^*$, $\pi_{1:2} = \pi_{1:2}^*$, and $v = v^*$ defined in (12), we set $M_{\lambda^*} = \|\lambda^*\|$ and obtain

$$\begin{aligned}&\sum_{t=1}^{K}\mathbb{E}\Big(\mathcal{L}(x_t,\lambda^*,\pi_{1:2}^*,v^*)-\mathcal{L}(x^*,\lambda_t,\pi_{1:2,t},v_t)\Big)+\frac{\alpha}{2}\mathbb{E}[\|\lambda_K-\lambda^*\|^2]\\&\le \frac{95KC_{f_1}^2C_{f_2}^2}{2\eta}+\sqrt{K}L_{f_1}\sigma_{f_2}^2+3\sqrt{K}C_{f_1}\sigma_{f_2}+\frac{\eta}{2}\|x_0-x^*\|^2\\&\quad+\frac{KD_X^2C_g^2}{\alpha}+\frac{10KC_g^2\|\lambda^*\|^2}{\eta}+\sum_{t=1}^{K}\frac{5C_g^2\mathbb{E}[\|\lambda_{t-1}\|^2]}{2\eta}+\sqrt{K}\|\lambda^*\|\sigma_g\\&\quad+\frac{\alpha}{2}\|\lambda_0-\lambda^*\|^2+\frac{K\sigma_g^2}{\alpha}+\sum_{t=1}^{K}\frac{5C_g^2\mathbb{E}[\|\lambda^*-\lambda_t\|^2]}{2\eta}.\end{aligned}$$

By using the fact that $\|\lambda_t\|^2 \leq 2\|\lambda_t - \lambda^*\|^2 + 2\|\lambda^*\|^2$, we further express the above inequality as

$$\begin{aligned}
&\sum_{t=1}^{K} \mathbb{E}\Big(\mathcal{L}(x_t, \lambda^*, \pi^*_{1:2}, v^*) - \mathcal{L}(x^*, \lambda_t, \pi_{1:2,t}, v_t)\Big) + \frac{\alpha}{2}\mathbb{E}[\|\lambda_K - \lambda^*\|^2] \\
&\quad \leq \frac{95KC_{f_1}^2 C_{f_2}^2}{2\eta} + \sqrt{K}L_{f_1}\sigma_{f_2}^2 + 3\sqrt{K}C_{f_1}\sigma_{f_2} + \frac{\eta}{2}\|x_0 - x^*\|^2 \\
&\quad\quad + \frac{KD_X^2 C_g^2}{\alpha} + \sum_{t=1}^{K} \frac{15C_g^2\|\lambda^*\|^2}{\eta} + \sum_{t=1}^{K} \frac{5C_g^2\mathbb{E}[\|\lambda_{t-1} - \lambda^*\|^2]}{\eta} + \sqrt{K}\|\lambda^*\|\sigma_g \\
&\quad\quad + \frac{\alpha}{2}\|\lambda_0 - \lambda^*\|^2 + \frac{K\sigma_g^2}{\alpha} + \sum_{t=1}^{K} \frac{5C_g^2\mathbb{E}[\|\lambda^* - \lambda_t\|^2]}{2\eta}.
\end{aligned}$$

By using the min-max relationship (12) that

$$\mathcal{L}(x_t, \lambda^*, \pi^*_{1:2}, v^*) - \mathcal{L}(x^*, \lambda_t, \pi_{1:2,t}, v_{1,t}) \geq 0,$$

dividing both sides of the above inequality by $\frac{\alpha}{2}$, and setting $\alpha = 2\sqrt{N}$ and $\eta = \frac{15C_g^2\sqrt{N}}{2}$, we obtain

$$\begin{aligned}
\mathbb{E}[\|\lambda_K - \lambda^*\|^2] &\leq \frac{2}{\alpha}\sum_{t=1}^{K} \mathbb{E}\Big(\mathcal{L}(x_t, \lambda^*, \pi^*_{1:2}, v^*) - \mathcal{L}(x^*, \lambda_t, \pi_{1:2,t}, v_t)\Big) + \mathbb{E}[\|\lambda_K - \lambda^*\|^2] \\
&\leq R_K + \frac{\mathbb{E}[\|\lambda^* - \lambda_K\|^2]}{3N} + \sum_{t=0}^{K-1} \frac{\mathbb{E}[\|\lambda^* - \lambda_t\|^2]}{N},
\end{aligned}$$

where

$$\begin{aligned}
R_K = {} & \frac{K}{N}\Big(\frac{7C_{f_1}^2 C_{f_2}^2}{C_g^2} + \frac{D_X^2 C_g^2}{2} + 2\|\lambda^*\|^2 + \frac{\sigma_g^2}{2}\Big) + \frac{15C_g^2}{4}\|x_0 - x^*\|^2 + \|\lambda_0 - \lambda^*\|^2 \\
& + \frac{\sqrt{K}}{\sqrt{N}}\Big(L_{f_1}\sigma_{f_2}^2 + 3C_{f_1}\sigma_{f_2} + \|\lambda^*\|\sigma_g\Big).
\end{aligned}$$

Next, letting R be the constant defined in (18), we note that $R_K \leq R$, and the above inequality further implies that

$$\Big(1 - \frac{1}{3N}\Big)\mathbb{E}[\|\lambda_K - \lambda^*\|^2] \leq R + \sum_{t=0}^{K-1} \frac{\mathbb{E}[\|\lambda^* - \lambda_t\|^2]}{N}.$$

Dividing both sides of the above inequality by $(1 - \frac{1}{3N})$, using the fact that $(1 - \frac{1}{3N}) \geq 1/2$ for $N \geq 2$, and applying Lemma 1, we conclude that

$$\begin{aligned}
\mathbb{E}[\|\lambda_K - \lambda^*\|^2] &\leq \Big(1 - \frac{1}{3N}\Big)^{-1}\Big(R + \sum_{t=0}^{K-1} \frac{\mathbb{E}[\|\lambda^* - \lambda_t\|^2]}{N}\Big) \\
&\leq 2R + 2\sum_{t=0}^{K-1} \frac{\mathbb{E}[\|\lambda^* - \lambda_t\|^2]}{N} \leq 2R\Big(1 + \frac{2}{N}\Big)^K \leq 2Re^2,
\end{aligned}$$

for all $1 \leq K \leq N$. This completes the proof. ∎

EC.1.7 Proof of Theorem 2

Proof: In our analysis, we set $\lambda_0 = 0$ for ease of presentation. Recall that $\bar{x}_N = \frac{1}{N}\sum_{t=1}^{N} x_t$, for any fixed $(\lambda, \pi_{1:2}, v)$, by the convexity of $\mathcal{L}(x, \lambda, \pi_{1:2}, v)$ with respect to x, we have

$$\frac{1}{N}\sum_{t=1}^{N} \mathcal{L}(x_t, \lambda, \pi_{1:2}, v) \geq \mathcal{L}(\bar{x}_N, \lambda, \pi_{1:2}, v). \tag{EC.19}$$

Meanwhile, we denote by

$$\hat{\pi}_2 \in \partial f_2(\bar{x}_N),\ \hat{\pi}_1 = \nabla f_1\Big(f_2(\bar{x}_N)\Big),\ \text{and}\ \ \hat{v} \in \partial g(\bar{x}_N), \tag{EC.20}$$

the dual variables associated with $\bar{x}_N$. By the definition of composite Lagrangian (10), we have

$$F(\bar{x}_N) = \mathcal{L}_F(\bar{x}_N, \hat{\pi}_{1:2}) = \mathcal{L}(\bar{x}_N, 0, \hat{\pi}_{1:2}, \hat{v}). \tag{EC.21}$$

First, we derive the convergence rate of the objective optimality gap $F(\bar{x}_N) - F(x^*)$. Let $\bar{\lambda}_N = \frac{1}{N}\sum_{t=1}^{N} \lambda_t$, $\bar{\pi}_{1:2,N} = \frac{1}{N}\sum_{t=1}^{N} \pi_{1:2,t}$, and $\bar{v}_N = \frac{1}{N}\sum_{t=1}^{N} v_t$. By setting $\lambda = 0$, $\pi_{1:2} = \hat{\pi}_{1:2}$, and $v = \hat{v}$ within (EC.20), and using (12) that $\mathcal{L}(x^*, \lambda^*, \pi^*_{1:2}, v^*) \geq \mathcal{L}(x^*, \lambda, \pi_{1:2}, v)$ for any $(\lambda, \pi_{1:2}, v) \in \mathbb{R}^m_+ \times \Pi \times \mathcal{V}$, we have

$$\begin{aligned}&\frac{1}{N}\sum_{t=1}^{N}\Big(\mathcal{L}(x_t, 0, \hat{\pi}_{1:2}, \hat{v}) - \mathcal{L}(x^*, \lambda_t, \pi_{1:2,t}, v_t)\Big)\\ &\qquad \geq \mathcal{L}(\bar{x}_N, 0, \hat{\pi}_{1:2}, \hat{v}) - \mathcal{L}(x^*, \lambda^*, \pi^*_{1:2}, v^*) = F(\bar{x}_N) - F(x^*),\end{aligned} \tag{EC.22}$$

where the last equality holds by (EC.21). Therefore, by (EC.22), setting the feasible point as $(x^*, \lambda, \pi_{1:2}, v) = (x^*, 0, \hat{\pi}_{1:2}, \hat{v})$, and applying Theorem 1 with the fact $\|\lambda\| = 0$ when $\lambda = 0$, we have

$$\begin{aligned}\mathbb{E}[F(\bar{x}_N)] - F(x^*) &\leq \frac{1}{N}\sum_{t=1}^{N}\mathbb{E}\Big(\mathcal{L}(x_t, 0, \hat{\pi}_{1:2}, \hat{v}) - \mathcal{L}(x^*, \lambda_t, \pi_{1:2,t}, v_t)\Big)\\ &\leq \frac{1}{\sqrt{N}}\Big(\frac{7C_{f_1}^2 C_{f_2}^2}{C_g^2} + L_{f_1}\sigma_{f_2}^2 + 3C_{f_1}\sigma_{f_2} + \frac{15C_g^2}{4}\|x_0 - x^*\|^2\Big)\\ &\quad + \frac{1}{N}\sum_{t=1}^{N}\Big(\frac{\mathbb{E}[\|\lambda_{t-1}\|^2]}{3\sqrt{N}} + \frac{\mathbb{E}[\|\lambda_t\|^2]}{3\sqrt{N}}\Big)\\ &\leq \frac{1}{\sqrt{N}}\Big(\frac{7C_{f_1}^2 C_{f_2}^2}{C_g^2} + L_{f_1}\sigma_{f_2}^2 + 3C_{f_1}\sigma_{f_2} + \frac{15C_g^2}{4}\|x_0 - x^*\|^2\Big) + \frac{(4\|\lambda^*\|^2 + 8Re^2)}{3\sqrt{N}},\end{aligned}$$

where the last inequality comes from the bound of $\|\lambda_t\|^2$ provided by Proposition 1 that

$$\mathbb{E}[\|\lambda_t\|^2] \leq 2\|\lambda^*\|^2 + 2\mathbb{E}[\|\lambda_t - \lambda^*\|^2] \leq 2\|\lambda^*\|^2 + 4Re^2,\ \text{for all}\ t = 1, 2, \cdots, N. \tag{EC.23}$$

This establishes the convergence rate for the objective optimality gap $F(\bar{x}_N) - F(x^*)$.

Second, we consider the feasibility residual $\|g(\bar{x}_N)_+\|_2$. Recall that $\hat{\pi}_{1:2}$ and $\hat{v}$ are the dual variables associated with $\bar{x}_N$ such that $\hat{\pi}_{1:2} \in \operatorname{argmax}_{\pi_{1,2}\in\Pi} \mathcal{L}_F(\bar{x}_N, \pi_{1:2})$ and $\hat{v} \in \operatorname{argmax}_{v\in\mathcal{V}} \mathcal{L}_g(\bar{x}_N, v)$, we have

$$F(\bar{x}_N) = \mathcal{L}_F(\bar{x}_N, \hat{\pi}_{1:2}) \geq \mathcal{L}_F(\bar{x}_N, \pi^*_{1:2})\ \text{and}\ g(\bar{x}_N) = \mathcal{L}_g(\bar{x}_N, \hat{v}) \geq \mathcal{L}_g(\bar{x}_N, v^*).$$

Since $\lambda^* \geq 0$, by the above inequalities and the min-max relationship (12), we can see

$$\begin{aligned}\mathcal{L}(\bar{x}_N, \lambda^*, \hat{\pi}_{1:2}, \hat{v}) &= \mathcal{L}_F(\bar{x}_N, \hat{\pi}) + \langle \lambda^*, \mathcal{L}_g(\bar{x}_N, \hat{v})\rangle \\ &\geq \mathcal{L}_F(\bar{x}_N, \pi^*_{1:2}) + \langle \lambda^*, \mathcal{L}_g(\bar{x}_N, v^*)\rangle = \mathcal{L}(\bar{x}_N, \lambda^*, \pi^*_{1:2}, v^*) \geq \mathcal{L}(x^*, \lambda^*, \pi^*_{1:2}, v^*),\end{aligned}$$

which further implies

$$F(\bar{x}_N) + \langle \lambda^*, g(\bar{x}_N)\rangle - F(x^*) \geq 0.$$

Meanwhile, due to the facts that $\lambda^* \geq 0$ and $g(\bar{x}_N) \leq g(\bar{x}_N)_+$, we have $\langle \lambda^*, g(\bar{x}_N)\rangle \leq \langle \lambda^*, g(\bar{x}_N)_+\rangle$, hence,

$$F(\bar{x}_N) + \|\lambda^*\|\|g(\bar{x}_N)_+\| - F(x^*) \geq F(\bar{x}_N) + \langle \lambda^*, g(\bar{x}_N)\rangle - F(x^*) \geq 0. \tag{EC.24}$$

Let $\tilde{\lambda} = (\|\lambda^*\|_2 + 1)g(\bar{x}_N)_+/\|g(\bar{x}_N)_+\|$. Consider another feasible point $(x^*, \tilde{\lambda}, \hat{\pi}_{1:2}, \hat{v})$, by (EC.19) and the facts that $g(\bar{x}_N)^\top g(\bar{x}_N)_+ = \|g(\bar{x}_N)_+\|^2$ and $\mathcal{L}(x^*, \lambda_t, \pi_{1:2,t}, v_t) \leq \mathcal{L}(x^*, \lambda^*, \pi^*_{1:2}, v^*)$, we have $\langle g(\bar{x}_N), \tilde{\lambda}\rangle = (\|\lambda^*\|_2 + 1)\|g(\bar{x}_N)_+\|$, which further yields that

$$\begin{aligned}&\frac{1}{N}\sum_{t=1}^{N}\Big(\mathcal{L}(x_t, \tilde{\lambda}, \hat{\pi}_{1:2}, \hat{v}) - \mathcal{L}(x^*, \lambda_t, \pi_{1:2,t}, v_t)\Big) \geq \mathcal{L}(\bar{x}_N, \tilde{\lambda}, \hat{\pi}_{1:2}, \hat{v}) - \mathcal{L}(x^*, \lambda^*, \pi^*_{1:2}, v^*) \\ &\quad = F(\bar{x}_N) + \langle \tilde{\lambda}, g(\bar{x}_N)\rangle - F(x^*) = F(\bar{x}_N) + (\|\lambda^*\|_2 + 1)\|g(\bar{x}_N)_+\| - F(x^*).\end{aligned}$$

By rearranging the terms in the inequality above and applying (EC.24), we obtain

$$\begin{aligned}\|g(\bar{x}_N)_+\| &\leq \frac{1}{N}\sum_{t=1}^{N}\Big(\mathcal{L}(x_t, \tilde{\lambda}, \hat{\pi}_{1:2}, \hat{v}) - (\mathcal{L}(x^*, \lambda_t, \pi_{1:2,t}, v_t)\Big) \\ &\quad - \Big(F(\bar{x}_N) + \|\lambda^*\|\|g(\bar{x}_N)_+\| - F(x^*)\Big) \\ &\leq \frac{1}{N}\sum_{t=1}^{N}\Big(\mathcal{L}(x_t, \tilde{\lambda}, \hat{\pi}_{1:2}, \hat{v}) - \mathcal{L}(x^*, \lambda_t, \pi_{1:2,t}, v_t)\Big).\end{aligned}$$

Next, we note that $\|\tilde{\lambda}\| = \|\lambda^*\| + 1$. By the inequality above, considering the feasible point $(x^*, \tilde{\lambda}, \hat{\pi}_{1:2}, \hat{v})$, and applying Theorem 1 with $\lambda_0 = 0$, $M_{\tilde{\lambda}} = \|\tilde{\lambda}\| = \|\lambda^*\| + 1$, $\alpha = 2\sqrt{N}$, and $\eta = \frac{15C_g^2\sqrt{N}}{2}$, we have

$$\begin{aligned}&\mathbb{E}[\|g(\bar{x}_N)_+\|] \\ &\leq \frac{95C_{f_1}^2C_{f_2}^2}{2\eta} + \frac{L_{f_1}\sigma_{f_2}^2}{\sqrt{N}} + \frac{3C_{f_1}\sigma_{f_2}}{\sqrt{N}} + \frac{\eta}{2N}\|x_0 - x^*\|^2 + \frac{D_X^2C_g^2}{\alpha} + \frac{\|\tilde{\lambda}\|\sigma_g}{\sqrt{N}} \\ &\quad + \frac{1}{N}\left[\frac{\alpha}{2}\|\tilde{\lambda}\|^2 + \frac{N\sigma_g^2}{\alpha} + \sum_{t=1}^{N}\left(\frac{5C_g^2\mathbb{E}[\|\lambda_{t-1}\|^2]}{2\eta} + \frac{10C_g^2\mathbb{E}[\|\tilde{\lambda}\|^2]}{\eta} + \frac{5C_g^2\mathbb{E}[\|\tilde{\lambda} - \lambda_t\|^2]}{2\eta}\right)\right] \\ &\leq \frac{1}{\sqrt{N}}\left(\frac{7C_{f_1}^2C_{f_2}^2}{C_g^2} + L_{f_1}\sigma_{f_2}^2 + 3C_{f_1}\sigma_{f_2} + \frac{15}{4}\|x_0 - x^*\|^2 + \frac{D_X^2C_g^2}{2} + \|\tilde{\lambda}\|\sigma_g + \|\tilde{\lambda}\|^2\right) \\ &\quad + \frac{\sigma_g^2}{2\sqrt{N}} + \frac{1}{N}\sum_{t=1}^{N}\left(\frac{\mathbb{E}[\|\lambda_{t-1}\|^2]}{3\sqrt{N}} + \frac{4\|\tilde{\lambda}\|^2}{3\sqrt{N}} + \frac{2(\|\tilde{\lambda}\|^2 + \mathbb{E}[\|\lambda_t\|^2])}{3\sqrt{N}}\right) \\ &\leq \frac{1}{\sqrt{N}}\left(\frac{7C_{f_1}^2C_{f_2}^2}{C_g^2} + L_{f_1}\sigma_{f_2}^2 + 3C_{f_1}\sigma_{f_2} + \frac{15C_g^2}{4}\|x_0 - x^*\|^2 + \frac{D_X^2C_g^2}{2} + \frac{\sigma_g^2}{2}\right) \\ &\quad + \frac{1}{\sqrt{N}}\Big((2\|\lambda^*\|^2 + 4Re^2) + 3\|\tilde{\lambda}\|^2 + \|\tilde{\lambda}\|\sigma_g\Big),\end{aligned}$$

where the second inequality holds because $\|\tilde{\lambda} - \lambda_t\|^2 \leq 2\|\tilde{\lambda}\|^2 + 2\|\lambda_t\|^2$, and the last inequality holds by the bound of $\|\lambda_t\|^2$ in (EC.23). The desired result can be acquired by substituting $\|\tilde{\lambda}\| = \|\lambda^*\|_2 + 1$ and $\|\tilde{\lambda}\|^2 = (\|\lambda^*\|_2 + 1)^2$ into the above inequality. This completes the proof. ∎

EC.2 Proof of Results in Section 3

EC.2.1 Proof of Lemma 3

First, we decompose $(\lambda - \lambda_t)^\top \mathcal{L}_G(x_t, v_{1:2,t})$ that

$$\begin{aligned}
&(\lambda - \lambda_t)^\top \mathcal{L}_G(x_t, v_{1:2,t}) \\
&\quad = (\lambda - \lambda_t)^\top \Big(\mathcal{L}_G(x_t, v_{1:2,t}) - \mathcal{L}_G(x_{t-1}, v_{1:2,t})\Big) + (\lambda - \lambda_t)^\top \mathcal{L}_G(x_{t-1}, v_{1:2,t}) \\
&\quad = \underbrace{\langle v_{2,t} v_{1,t}(\lambda - \lambda_t), x_t - x_{t-1}\rangle}_{T_{1,t}} + \underbrace{(\lambda - \lambda_t)^\top [\mathcal{L}_G(x_{t-1}, v_{1:2,t}) - H_t]}_{T_{2,t}} + \underbrace{(\lambda - \lambda_t)^\top H_t}_{T_{3,t}}.
\end{aligned} \tag{EC.25}$$

We now provide bounds for the three terms $T_{1,t}, T_{2,t}$, and $T_{3,t}$. First, for $T_{1,t}$, by the fact that $\langle a, b\rangle \leq \frac{5\|a\|^2}{2\eta} + \frac{\eta\|b\|^2}{10}$ for any vectors a, b, we have

$$\begin{aligned}
\sum_{t=1}^K \mathbb{E}[T_{1,t}] &= \sum_{t=1}^K \mathbb{E}[(x_t - x_{t-1})^\top v_{2,t} v_{1,t}(\lambda - \lambda_t)] \\
&\leq \sum_{t=1}^K \frac{5\mathbb{E}[\|v_{2,t} v_{1,t}(\lambda - \lambda_t)\|^2]}{2\eta} + \sum_{t=1}^K \frac{\eta \mathbb{E}[\|x_t - x_{t-1}\|^2]}{10} \\
&\leq \sum_{t=1}^K \frac{5C_{g_1}^2 C_{g_2}^2 \mathbb{E}[\|\lambda - \lambda_t\|^2]}{2\eta} + \sum_{t=1}^K \frac{\eta \mathbb{E}[\|x_t - x_{t-1}\|^2]}{10},
\end{aligned} \tag{EC.26}$$

where the last inequality holds by Assumption 4 (d)-(e) that $\|v_{1,t}\|^2 \leq C_{g_1}^2$ and $\|v_{2,t}\|^2 \leq C_{g_2}^2$. Second, we consider $T_{2,t}$ and denote by $\tilde{\Delta}_{2,t} = \mathcal{L}_G(x_{t-1}, v_{1:2,t}) - H_t$. Given x_{t-1}, we observe from Algorithm 2 that H_t is conditionally independent of λ_{t-1}. Together with Lemma 2(a) that H_t is also an unbiased estimator for $\mathcal{L}_G(x_{t-1}, v_{1:2,t})$, we have $\mathbb{E}[\lambda_{t-1}^\top \tilde{\Delta}_{2,t}] = 0$, which further implies that

$$\begin{aligned}
\mathbb{E}[T_{2,t}] &= \mathbb{E}[\lambda^\top \tilde{\Delta}_{2,t}] - \mathbb{E}[\tilde{\Delta}_{2,t}^\top \lambda_{t-1}] + \mathbb{E}[(\lambda_{t-1} - \lambda_t)^\top \tilde{\Delta}_{2,t}] \\
&= \mathbb{E}[\lambda^\top \tilde{\Delta}_{2,t}] + \mathbb{E}[(\lambda_{t-1} - \lambda_t)^\top \tilde{\Delta}_{2,t}].
\end{aligned}$$

By setting $M_\lambda = \|\lambda\|$ and recalling Lemma 2(b) that $\mathrm{Var}(\tilde{\Delta}_{2,t}) \leq \sigma_H^2$, we further have

$$\begin{aligned}
\sum_{t=1}^K \mathbb{E}[T_{2,t}] &= \sum_{t=1}^K \mathbb{E}\Big[\lambda^\top \tilde{\Delta}_{2,t}\Big] + \sum_{t=1}^K \mathbb{E}\Big[(\lambda_{t-1} - \lambda_t)^\top \tilde{\Delta}_{2,t}\Big] \\
&\leq \sqrt{K} M_\lambda \sigma_H + \sum_{t=1}^K \mathbb{E}\Big[(\lambda_{t-1} - \lambda_t)^\top \tilde{\Delta}_{2,t}\Big] \\
&\leq \sqrt{K} M_\lambda \sigma_H + \sum_{t=1}^K \left(\frac{\alpha \mathbb{E}[\|\lambda_{t-1} - \lambda_t\|^2]}{4} + \frac{\sigma_H^2}{\alpha}\right),
\end{aligned} \tag{EC.27}$$

where the first inequality holds by Lemma EC.2 in Appendix Section EC.1, and the last inequality holds by the fact $\langle a, b\rangle \leq \frac{\|a\|^2}{4} + \|b\|^2$ for any vectors a, b.

Finally, for $T_{3,t}$, by applying the three-point Lemma EC.1 to the update rule that

$$\lambda_t \in \operatorname*{argmax}_{\lambda \in \mathbb{R}_+^m} \Big\{ \lambda^\top H_t - \frac{\alpha}{2}\|\lambda_{t-1} - \lambda\|^2 \Big\},$$

we have that for any $\lambda \in \mathbb{R}_+^m$,

$$(\lambda - \lambda_t)^\top H_t = -(\lambda_t - \lambda)^\top H_t \le \frac{\alpha}{2}\|\lambda_{t-1} - \lambda\|^2 - \frac{\alpha}{2}\|\lambda_{t-1} - \lambda_t\|^2 - \frac{\alpha}{2}\|\lambda_t - \lambda\|^2,$$

which implies

$$\mathbb{E}[T_{3,t}] \le \sum_{t=1}^{K} \Big(\frac{\alpha}{2}\|\lambda_{t-1} - \lambda\|^2 - \frac{\alpha}{2}\|\lambda_{t-1} - \lambda_t\|^2 - \frac{\alpha}{2}\|\lambda_t - \lambda\|^2 \Big). \tag{EC.28}$$

Summing (EC.25) over $t = 1, \cdots, K$, taking expectations, and plugging (EC.26), (EC.27), and (EC.28) in, we obtain that for all $\lambda \in \mathbb{R}_+^m$,

$$\begin{aligned}
&\mathbb{E}\Big[\sum_{t=1}^{K} (\lambda - \lambda_t)^\top \mathcal{L}_G(x_t, v_{1,t}, v_{2,t}) \Big] \\
&\le \frac{\alpha}{2}\mathbb{E}[\|\lambda_0 - \lambda\|^2] - \frac{\alpha}{2}\mathbb{E}[\|\lambda_K - \lambda\|^2] - \sum_{t=1}^{K} \frac{\alpha}{2}\mathbb{E}[\|\lambda_{t-1} - \lambda_t\|^2] + \sqrt{K} M_\lambda \sigma_H \\
&\quad + \sum_{t=1}^{K} \Big(\frac{\alpha \mathbb{E}[\|\lambda_{t-1} - \lambda_t\|^2]}{4} + \frac{\sigma_H^2}{\alpha} + \frac{5 C_{g_1}^2 C_{g_2}^2 \mathbb{E}[\|\lambda - \lambda_t\|^2]}{2\eta} + \frac{\eta \mathbb{E}[\|x_t - x_{t-1}\|^2]}{10} \Big) \\
&\le \frac{\alpha}{2}\|\lambda_0 - \lambda\|^2 - \frac{\alpha}{2}\mathbb{E}[\|\lambda_K - \lambda\|^2] + \sqrt{K} M_\lambda \sigma_H \\
&\quad + \sum_{t=1}^{K} \Big(\frac{\sigma_H^2}{\alpha} - \frac{\alpha \mathbb{E}[\|\lambda_{t-1} - \lambda_t\|^2]}{4} + \frac{5 C_{g_1}^2 C_{g_2}^2 \mathbb{E}[\|\lambda - \lambda_t\|^2]}{2\eta} + \frac{\eta \mathbb{E}[\|x_t - x_{t-1}\|^2]}{10} \Big),
\end{aligned}$$

which completes the proof. ∎

EC.2.2 Proof of Theorem 3

We provide bound on each term after decomposing (28). Consider any integer K such that $K \le N$. First, by Lemma EC.7 in Appendix Section EC.2, we obtain

$$\begin{aligned}
&\sum_{t=1}^{K} \mathbb{E}\Big[\lambda^\top \Big(\mathcal{L}_G(x_t, v_{1:2}) - \mathcal{L}_G(x_t, v_{1:2,t}) \Big) \Big] \\
&\quad \le \sum_{t=1}^{K} \frac{\eta}{10}\mathbb{E}\Big[\|x_t - x_{t-1}\|^2 \Big] + \frac{40 K C_{g_1}^2 C_{g_2}^2 M_\lambda^2}{\eta} + 2\sqrt{K} L_{g_1} \sigma_{g_2}^2 M_\lambda + 3\sqrt{K} C_{g_1} \sigma_{g_2} M_\lambda.
\end{aligned} \tag{EC.29}$$

Second, by Lemma EC.8 in Appendix Section EC.2, we have

$$\begin{aligned}
&\sum_{t=1}^{K} \mathbb{E}\Big[\mathcal{L}_F(x_t, \pi_{1:2,t}) - \mathcal{L}_F(x^*, \pi_{1:2,t}) + \lambda_t^\top \Big(\mathcal{L}_G(x_t, v_{1:2,t}) - \mathcal{L}_G(x^*, v_{1:2,t}) \Big) \Big] \\
&\quad \le \frac{\eta}{2}\|x_0 - x^*\|^2 - \sum_{t=1}^{K} \frac{3\eta}{10}\mathbb{E}\Big[\|x_{t-1} - x_t\|^2 \Big] + \sum_{t=1}^{K} \frac{5 C_{f_1}^2 C_{f_2}^2}{2\eta} \\
&\qquad + \sum_{t=1}^{K} \frac{\alpha}{4}\mathbb{E}\Big[\|\lambda_t - \lambda_{t-1}\|^2 \Big] + \sum_{t=1}^{K} \frac{D_X^2 C_{g_1}^2 C_{g_2}^2}{\alpha} + \sum_{t=1}^{K} \frac{5 C_{g_1}^2 C_{g_2}^2 \mathbb{E}[\|\lambda_{t-1}\|^2]}{2\eta}.
\end{aligned} \tag{EC.30}$$

Meanwhile, Lemma EC.3 in Appendix Section EC.1.2 implies that

$$\begin{aligned}
&\sum_{t=1}^{K}\mathbb{E}\Big[\mathcal{L}_F(x_t,\pi_{1:2})-\mathcal{L}_F(x_t,\pi_{1:2,t})\Big] \\
&\quad\le \sum_{t=1}^{K}\frac{\eta}{10}\mathbb{E}\Big[\|x_t-x_{t-1}\|^2\Big]+\frac{40KC_{f_1}^2C_{f_2}^2}{\eta}+2\sqrt{K}L_{f_1}\sigma_{f_2}^2+3\sqrt{K}C_{f_1}\sigma_{f_2}.
\end{aligned} \tag{EC.31}$$

Finally, by Lemma 3 and the assumption that $\|\lambda\|\le M_\lambda$, we have

$$\begin{aligned}
&\mathbb{E}\Big[\sum_{t=1}^{K}(\lambda-\lambda_t)^\top\mathcal{L}_G(x_t,v_{1:2,t})\Big]+\frac{\alpha}{2}\mathbb{E}\Big[\|\lambda_K-\lambda\|^2\Big] \\
&\le \sum_{t=1}^{K}\Big(\frac{\sigma_H^2}{\alpha}-\frac{\alpha\mathbb{E}[\|\lambda_{t-1}-\lambda_t\|^2]}{4}+\frac{5C_{g_1}^2C_{g_2}^2\mathbb{E}[\|\lambda-\lambda_t\|^2]}{2\eta}+\frac{\eta\mathbb{E}[\|x_t-x_{t-1}\|^2]}{10}\Big) \\
&\quad+\frac{\alpha}{2}\mathbb{E}\Big[\|\lambda_0-\lambda\|^2\Big]+\sqrt{K}M_\lambda\sigma_H.
\end{aligned} \tag{EC.32}$$

By substituting (EC.29), (EC.30), (EC.31), and (EC.32) into (28), and omitting some algebras, we obtain

$$\begin{aligned}
&\sum_{t=1}^{K}\mathbb{E}\Big(\mathcal{L}(x_t,\lambda,\pi_{1:2},v_{1:2})-\mathcal{L}(x^*,\lambda_t,\pi_{1:2,t},v_{1:2,t})\Big)+\frac{\alpha}{2}\mathbb{E}[\|\lambda_K-\lambda\|^2] \\
&\quad\le \frac{40KC_{g_1}^2C_{g_2}^2M_\lambda^2}{\eta}+2\sqrt{K}L_{g_1}\sigma_{g_2}^2M_\lambda+3\sqrt{K}C_{g_1}\sigma_{g_2}M_\lambda+\frac{\eta}{2}\|x_0-x^*\|^2 \\
&\quad\quad+\frac{95KC_{f_1}^2C_{f_2}^2}{2\eta}+\frac{K}{\alpha}(D_X^2C_{g_1}^2C_{g_2}^2+\sigma_H^2)+\frac{\alpha}{2}\mathbb{E}[\|\lambda_0-\lambda\|^2]+\sqrt{K}M_\lambda\sigma_H \\
&\quad\quad+2\sqrt{K}L_{f_1}\sigma_{f_2}^2+3\sqrt{K}C_{f_1}\sigma_{f_2}+\sum_{t=1}^{K}\frac{5C_{g_1}^2C_{g_2}^2}{2\eta}\Big(\mathbb{E}[\|\lambda-\lambda_t\|^2]+\mathbb{E}[\|\lambda_{t-1}\|^2]\Big).
\end{aligned}$$

This completes the proof. ∎

Lemma EC.7 *Suppose Assumptions 1, 3, and 4 hold, and Algorithm 2 generates* $\{(x_t,\lambda_t,\pi_{1:2,t},v_{1:2,t})\}_{t=1}^N$ *by setting* $\tau_t=\rho_t=t/2$, $\eta_t=\eta$, *and* $\alpha_t=\alpha$ *for* $t=1,2,\cdots,N$. *Letting* $\lambda\in\mathbb{R}_+^m$ *be a bounded random variable satisfying* $\|\lambda\|\le M_\lambda$ *uniformly, then for any* $K\le N$, *we have*

$$\begin{aligned}
&\sum_{t=1}^{K}\mathbb{E}\Big[\lambda^\top\Big(\mathcal{L}_G(x_t,v_{1:2})-\mathcal{L}_G(x_t,v_{1:2,t})\Big)\Big] \\
&\quad\le \sum_{t=1}^{K}\frac{\eta}{10}\mathbb{E}[\|x_t-x_{t-1}\|^2]+\frac{60KC_{g_1}^2C_{g_2}^2M_\lambda^2}{\eta}+2\sqrt{K}L_{g_1}M_\lambda\sigma_{g_2}^2+4\sqrt{K}M_\lambda C_{g_1}\sigma_{g_2}.
\end{aligned}$$

Proof: This result is similar to Lemma EC.3 with an additional nonnegative (random) variable λ. We start with the following decomposition that

$$\begin{aligned}
&\lambda^\top(\mathcal{L}_G(x_t,v_{1:2})-\mathcal{L}_G(x_t,v_{1:2,t})) \\
&\quad=\lambda^\top(\mathcal{L}_{g_1}(x_t,v_1,v_2)-\mathcal{L}_{g_1}(x_t,v_1,v_{2,t}))+\lambda^\top(\mathcal{L}_{g_1}(x_t,v_1,v_{2,t})-\mathcal{L}_{g_1}(x_t,v_{1,t},v_{2,t})).
\end{aligned}$$

First, we recall (25) that $v_{2,t} \in \mathrm{argmax}_{v_2 \in \mathcal{V}_2}\{v_2^\top x_{t-1} - g_2^*(v_2)\}$. By Assumption 4 (g) that for each compositional constraint $G^{(j)} = g_1^{(j)} \circ g_2^{(j)}(x)$, the outer-level function $g_1^{(j)}$ is monotone nondecreasing for non-affine inner-level function $g_2^{(j)}$, we obtain

$$\mathbb{E}\Big[\Big\langle (v_2 - v_{2,t})^\top x_{t-1} - g_2^*(v_2) + g_2^*(v_{2,t}), v_1 \Big\rangle\Big] \le 0.$$

Because $\lambda \ge 0$, by following (EC.6) in Lemma EC.3, we have

$$\begin{aligned}
\sum_{t=1}^K \mathbb{E}\Big[\lambda^\top\Big(\mathcal{L}_{g_1}(x_t, v_1, v_2) - \mathcal{L}_{g_1}(x_t, v_1, v_{2,t})\Big)\Big] &\le \sum_{t=1}^K \mathbb{E}[\langle (v_2 - v_{2,t})v_1\lambda, x_t - x_{t-1}\rangle] \\
&\le \sum_{t=1}^K \Big[\frac{10\mathbb{E}[\|(v_2 - v_{2,t})v_1\lambda\|^2]}{\eta} + \frac{\eta}{20}\mathbb{E}[\|x_t - x_{t-1}\|^2]\Big] \\
&\le \sum_{t=1}^K \Big[\frac{40C_{g_1}^2 C_{g_2}^2 M_\lambda^2}{\eta} + \frac{\eta}{20}\mathbb{E}[\|x_t - x_{t-1}\|^2]\Big],
\end{aligned}$$

where the last inequality uses the fact that $\|\lambda\| \le M_\lambda$. Second, by Proposition 15 in **?**, the L_{g_1} smoothness of g_1 implies $\lambda^\top D_{g_1^*}(v_1;\bar{v}_1) \ge \frac{\|\lambda^\top(v_1 - \bar{v}_1)\|^2}{2\|\lambda\|L_{g_1}} \ge \frac{\|\lambda^\top(v_1-\bar{v}_1)\|^2}{2M_\lambda L_{g_1}}$, which further yields

$$\begin{aligned}
&\sum_{t=1}^K \mathbb{E}\Big[\lambda^\top(v_{1,t} - v_{1,t-1})^\top(g_2(x_{t-1}, \zeta_{2,t-1}^1) - g_2(x_{t-1})) - \rho_t\lambda^\top D_{g_1^*}(v_{1,t}; v_{1,t-1})\Big] \\
&\quad\le \sum_{t=2}^K \frac{M_\lambda L_{g_1}\sigma_{g_2}^2}{\rho_t} + \sqrt{\mathbb{E}[\|\lambda\|^2\|v_{1,1} - v_{1,0}\|^2]}\sqrt{\mathbb{E}\Big[\Big\|g_2(x_0, \zeta_{2,0}^1) - g_2(x_0)\Big\|^2\Big]} \\
&\quad\le 2L_{g_1}M_\lambda\sigma_{g_2}^2\log K + 2M_\lambda C_{g_1}\sigma_{g_2} \\
&\quad\le 2\sqrt{K}L_{g_1}M_\lambda\sigma_{g_2}^2 + 2\sqrt{K}M_\lambda C_{g_1}\sigma_{g_2}.
\end{aligned}$$

By substituting the above bound into Proposition 6 of **?**, we further obtain that

$$\begin{aligned}
&\sum_{t=1}^K \mathbb{E}\Big[\lambda^\top\Big(\mathcal{L}_{g_1}(x_t, v_1, v_{2,t}) - \mathcal{L}_{g_1}(x_t, v_{1,t}, v_{2,t})\Big)\Big] \\
&\quad\le \sum_{t=1}^K \Big[\frac{20C_{g_1}^2 C_{g_2}^2 M_\lambda^2}{\eta} + \frac{\eta}{20}\mathbb{E}[\|x_t - x_{t-1}\|^2]\Big] + 2\sqrt{K}L_{g_1}M_\lambda\sigma_{g_2}^2 + 4\sqrt{K}M_\lambda C_{g_1}\sigma_{g_2}.
\end{aligned}$$

The desired result then follows by combining the preceding inequalities. ∎

Lemma EC.8 *Suppose Assumptions 1, 3, and 4 hold, and Algorithm 2 generates* $\{(x_t, \lambda_t, \pi_t, v_t)\}_{t=1}^N$ *by setting* $\rho_t = \tau_t = \frac{t}{2}$, $\eta_t = \eta$, *and* $\alpha_t = \alpha$ *for* $t \le N$. *Then for any integer* $K \le N$, *we have*

$$\begin{aligned}
&\sum_{t=1}^K \mathbb{E}\Big[\mathcal{L}_F(x_t, \pi_{1:2,t}) - \mathcal{L}_F(x^*, \pi_{1:2,t}) + \lambda_t^\top\Big(\mathcal{L}_G(x_t, v_{1:2,t}) - \mathcal{L}_G(x^*, v_{1:2,t})\Big)\Big] \\
&\quad\le \frac{\eta}{2}\|x_0 - x^*\|^2 - \sum_{t=1}^K \frac{3\eta_{t-1}}{10}\mathbb{E}\Big[\|x_{t-1} - x_t\|^2\Big] + \sum_{t=1}^K \frac{5C_{f_1}^2 C_{f_2}^2}{2\eta} \\
&\qquad + \sum_{t=1}^K \frac{\alpha}{4}\mathbb{E}[\|\lambda_t - \lambda_{t-1}\|^2] + \sum_{t=1}^K \frac{D_X^2 C_{g_1}^2 C_{g_2}^2}{\alpha} + \sum_{t=1}^K \frac{5C_{g_1}^2 C_{g_2}^2\mathbb{E}[\|\lambda_{t-1}\|^2]}{2\eta}.
\end{aligned}$$

Proof: Recall the update rule for x_t (26) that

$$x_t = \operatorname*{argmin}_{x\in\mathcal{X}} \Big\{ \Big\langle \pi^0_{2,t}\pi^0_{1,t} + v^0_{2,t}v^0_{1,t}\lambda_{t-1}, x \Big\rangle + \frac{\eta}{2}\|x_{t-1}-x\|^2 \Big\}.$$

By following analogous analysis in Lemma EC.4, and applying the three-point Lemma EC.1 to the above update rule, we have

$$\begin{aligned}
&\mathcal{L}_F(x_t,\pi_{1:2,t}) - \mathcal{L}_F(x^*,\pi_{1:2,t}) + \lambda_t^\top\Big(\mathcal{L}_G(x_t,v_{1:2,t}) - \mathcal{L}_G(x^*,v_{1:2,t})\Big)\\
&= (x_t-x^*)^\top \pi_{2,t}\pi_{1,t} + (x_t-x^*)^\top v_{2,t}v_{1,t}\lambda_t\\
&\le \frac{\eta}{2}\|x_{t-1}-x^*\|^2 - \frac{\eta}{2}\|x_{t-1}-x_t\|^2 - \frac{\eta}{2}\|x_t-x^*\|\\
&\quad + \langle \pi_{2,t}\pi_{1,t} - \pi^0_{2,t}\pi^0_{1,t}, x_t-x^*\rangle + \tilde\Lambda_{1,t} + \tilde\Lambda_{2,t},
\end{aligned} \tag{EC.33}$$

where

$$\tilde\Lambda_{1,t} = (x_t-x^*)^\top v_{2,t}v_{1,t}(\lambda_t-\lambda_{t-1}) \text{ and } \tilde\Lambda_{2,t} = (x_t-x^*)^\top\Big(v_{2,t}v_{1,t} - v^0_{2,t}v^0_{1,t}\Big)\lambda_{t-1}.$$

We note that a bound for $\Big\langle \pi_{2,t}\pi_{1,t} - \pi^0_{2,t}\pi^0_{1,t}, x_t-x^*\Big\rangle$ has been provided in (EC.9). Specifically, we have

$$\begin{aligned}
\mathbb{E}\Big[\sum_{t=1}^K \Big\langle \pi_{2,t}\pi_{1,t} - \pi^0_{2,t}\pi^0_{1,t}, x_t-x^*\Big\rangle\Big] &= \mathbb{E}\Big[\sum_{t=1}^K \Big\langle \pi_{2,t}\pi_{1,t} - \pi^0_{2,t}\pi^0_{1,t}, x_t-x_{t-1}\Big\rangle\Big]\\
&\le \sum_{t=1}^K \Big(\frac{5C^2_{f_1}C^2_{f_2}}{2\eta} + \frac{\eta\mathbb{E}[\|x_t-x_{t-1}\|^2]}{10}\Big).
\end{aligned} \tag{EC.34}$$

To bound $\tilde\Lambda_{1,t}$, by following the analysis of (EC.10), we have

$$\begin{aligned}
\mathbb{E}[\tilde\Lambda_{1,t}] &\le \frac{\alpha}{4}\mathbb{E}[\|\lambda_t-\lambda_{t-1}\|^2] + \frac{1}{\alpha}\mathbb{E}[\|(x_t-x^*)^\top v_{2,t}v_{1,t}\|^2]\\
&\le \frac{\alpha}{4}\mathbb{E}[\|\lambda_t-\lambda_{t-1}\|^2] + \frac{D^2_X C^2_{g_1}C^2_{g_2}}{\alpha}.
\end{aligned} \tag{EC.35}$$

Meanwhile, we observe from Algorithm 2 that $\mathbb{E}[x^\top_{t-1}(v^0_{2,t}v^0_{1,t} - v_{2,t}v_{1,t})\lambda_{t-1}] = 0$. By following the analysis of (EC.11), we have

$$\begin{aligned}
\mathbb{E}[\sum_{t=1}^K \tilde\Lambda_{2,t}] &= \sum_{t=1}^K \mathbb{E}[(x_t-x_{t-1})^\top(v_{2,t}v_{1,t} - v^0_{2,t}v^0_{1,t})\lambda_{t-1}]\\
&\le \sum_{t=1}^K \mathbb{E}[\frac{5\|(v^0_{2,t}v^0_{1,t} - v_{2,t}v_{1,t})\lambda_{t-1}\|^2}{2\eta}] + \sum_{t=1}^K \frac{\eta\mathbb{E}[\|x_t-x_{t-1}\|^2]}{10}\\
&\le \sum_{t=1}^K \frac{5C^2_{g_1}C^2_{g_2}\mathbb{E}[\|\lambda_{t-1}\|^2]}{2\eta} + \sum_{t=1}^K \frac{\eta\mathbb{E}[\|x_t-x_{t-1}\|^2]}{10},
\end{aligned} \tag{EC.36}$$

where the last inequality uses the fact that λ_{t-1} and $v^0_{2,t}v^0_{1,t} - v_{2,t}v_{1,t}$ are independent. Summing

(EC.33) over $t = 1, 2, \cdots, K$, and combining (EC.34), (EC.35), and (EC.36), we obtain

$$
\begin{aligned}
&\sum_{t=1}^{K} \mathbb{E}\Big[\mathcal{L}_F(x_t, \pi_{1,t}, \pi_{2,t}) - \mathcal{L}_F(x^*, \pi_{1,t}, \pi_{2,t}) + \lambda_t^\top \Big(\mathcal{L}_G(x_t, v_t) - \mathcal{L}_G(x^*, v_t)\Big)\Big] \\
&\quad \le \frac{\eta}{2}\|x_0 - x^*\|^2 + \sum_{t=1}^{K}\left(\frac{\eta}{10} + \frac{\eta}{10} - \frac{\eta}{2}\right)\mathbb{E}\Big[\|x_{t-1} - x_t\|^2\Big] + \sum_{t=1}^{K} \frac{5C_{f_1}^2 C_{f_2}^2}{2\eta} \\
&\qquad + \sum_{t=1}^{K} \frac{\alpha}{4}\mathbb{E}[\|\lambda_t - \lambda_{t-1}\|^2] + \sum_{t=1}^{K} \frac{D_X^2 C_{g_1}^2 C_{g_2}^2}{\alpha} + \sum_{t=1}^{K} \frac{5C_{g_1}^2 C_{g_2}^2 \mathbb{E}[\|\lambda_{t-1}\|^2]}{2\eta} \\
&\quad = \frac{\eta}{2}\|x_0 - x^*\|^2 - \sum_{t=1}^{K} \frac{3\eta}{10}\mathbb{E}\Big[\|x_{t-1} - x_t\|^2\Big] + \sum_{t=1}^{K} \frac{5C_{f_1}^2 C_{f_2}^2}{2\eta} \\
&\qquad + \sum_{t=1}^{K} \frac{\alpha}{4}\mathbb{E}[\|\lambda_t - \lambda_{t-1}\|^2] + \sum_{t=1}^{K} \frac{D_X^2 C_{g_1}^2 C_{g_2}^2}{\alpha} + \sum_{t=1}^{K} \frac{5C_{g_1}^2 C_{g_2}^2 \mathbb{E}[\|\lambda_{t-1}\|^2]}{2\eta}.
\end{aligned}
$$

This completes the proof. ∎

EC.2.3 Proof of Proposition 2

For the feasible point $(x^*, \lambda, \pi_{1:2}, v_{1:2})$, we set $\pi_{1:2} = \pi_{1:2}^*$, $v_{1:2} = v_{1:2}^*$, and $\lambda = \lambda^*$ as defined in (24). By applying Theorem 3 with $M_{\lambda^*} = \|\lambda^*\|$, we obtain

$$
\begin{aligned}
&\sum_{t=1}^{K} \mathbb{E}\Big(\mathcal{L}(x_t, \lambda^*, \pi_{1:2}^*, v_{1:2}^*) - \mathcal{L}(x^*, \lambda_t, \pi_{1:2,t}, v_{1:2,t})\Big) + \frac{\alpha}{2}\mathbb{E}[\|\lambda_K - \lambda^*\|^2] \\
&\le \frac{40KC_{g_1}^2 C_{g_2}^2\|\lambda^*\|}{\eta} + 2\sqrt{K}L_{g_1}\sigma_{g_2}^2\|\lambda^*\| + 3\sqrt{K}C_{g_1}\sigma_{g_2}\|\lambda^*\| + \frac{\eta}{2}\|x_0 - x^*\|^2 \\
&\quad + \frac{95KC_{f_1}^2 C_{f_2}^2}{2\eta} + \frac{K}{\alpha}(D_X^2 C_{g_1}^2 C_{g_2}^2 + \sigma_H^2) + \frac{\alpha}{2}\|\lambda_0 - \lambda^*\|^2 + \sqrt{K}\|\lambda^*\|\sigma_H \\
&\quad + 2\sqrt{K}L_{f_1}\sigma_{f_2}^2 + 3\sqrt{K}C_{f_1}\sigma_{f_2} + \sum_{t=1}^{K} \frac{5C_{g_1}^2 C_{g_2}^2}{2\eta}\Big(\mathbb{E}[\|\lambda^* - \lambda_t\|^2] + \mathbb{E}[\|\lambda_{t-1}\|^2]\Big).
\end{aligned}
$$

We note that $\mathcal{L}(x_t, \lambda^*, \pi_{1:2}^*, v_{1:2}^*) - \mathcal{L}(x^*, \lambda_t, \pi_{1:2,t}, v_{1:2,t}) \ge 0$ by the definition of a saddle point (24). Thus, we have

$$
\begin{aligned}
&\frac{\alpha}{2}\mathbb{E}[\|\lambda_K - \lambda^*\|^2] \\
&\le \frac{40KC_{g_1}^2 C_{g_2}^2\|\lambda^*\|}{\eta} + \frac{\eta}{2}\|x_0 - x^*\|^2 + \frac{95KC_{f_1}^2 C_{f_2}^2}{2\eta} + \frac{K}{\alpha}(D_X^2 C_{g_1}^2 C_{g_2}^2 + \sigma_H^2) \\
&\quad + \frac{\alpha}{2}\|\lambda_0 - \lambda^*\|^2 + \sum_{t=1}^{K} \frac{5C_{g_1}^2 C_{g_2}^2}{2\eta}\Big(\mathbb{E}[\|\lambda^* - \lambda_t\|^2] + \mathbb{E}[\|\lambda_{t-1}\|^2]\Big) + \sqrt{K}Q,
\end{aligned}
$$

where $Q = 2L_{g_1}\sigma_{g_2}^2\|\lambda^*\| + 3C_{g_1}\sigma_{g_2}\|\lambda^*\| + \|\lambda^*\|\sigma_H + 2L_{f_1}\sigma_{f_2}^2 + 3C_{f_1}\sigma_{f_2}$ is defined in (29). By using the fact that $\|\lambda_t\|^2 \le 2\|\lambda_t - \lambda^*\|^2 + 2\|\lambda^*\|^2$, setting $\alpha = 2\sqrt{N}$ and $\eta = \frac{15C_{g_1}^2 C_{g_2}^2\sqrt{N}}{2}$, and dividing both sides of the above inequality by $\sqrt{N}$, we obtain

$$
\begin{aligned}
&\mathbb{E}[\|\lambda_K - \lambda^*\|^2] \\
&\leq \frac{16K\|\lambda^*\|}{3N} + \frac{15C_{g_1}^2C_{g_2}^2}{4}\|x_0 - x^*\|^2 + \frac{19KC_{f_1}^2C_{f_2}^2}{3C_{g_1}^2C_{g_2}^2N} + \frac{K}{2N}(D_X^2C_{g_1}^2C_{g_2}^2 + \sigma_H^2) \\
&\quad + \|\lambda_0 - \lambda^*\|^2 + \frac{1}{3N}\sum_{t=1}^{K}\Big(\mathbb{E}[\|\lambda^* - \lambda_t\|^2] + 2\mathbb{E}[\|\lambda_{t-1} - \lambda^*\|^2] + 2\|\lambda^*\|^2\Big) + \frac{\sqrt{K}Q}{\sqrt{N}} \\
&= R_K + \frac{1}{3N}\mathbb{E}[\|\lambda^* - \lambda_K\|^2] + \frac{1}{N}\sum_{t=1}^{K}\mathbb{E}[\|\lambda_{t-1} - \lambda^*\|^2],
\end{aligned}
$$

where

$$
R_K = \frac{K}{N}\left(6\|\lambda^*\| + \frac{7C_{f_1}^2C_{f_2}^2}{C_{g_1}^2C_{g_2}^2} + \frac{D_X^2C_{g_1}^2C_{g_2}^2 + \sigma_H^2}{2}\right) + \frac{15C_{g_1}^2C_{g_2}^2}{4}\|x_0 - x^*\|^2 + \|\lambda_0 - \lambda^*\|^2 + \frac{\sqrt{K}Q}{\sqrt{N}}.
$$

By noting that $R_K \leq R$ defined in (29), the above inequality further implies that

$$
\left(1 - \frac{1}{3N}\right)\mathbb{E}[\|\lambda_K - \lambda^*\|^2] \leq R + \sum_{t=1}^{K}\frac{\mathbb{E}[\|\lambda_{t-1} - \lambda^*\|^2]}{N}.
$$

Assuming $N \geq 2$, for all $K \leq N$, we have

$$
\mathbb{E}[\|\lambda_K - \lambda^*\|^2] \leq 2R + \sum_{t=0}^{K-1}\frac{2}{N}\cdot\mathbb{E}[\|\lambda_t - \lambda^*\|^2].
$$

Finally, by applying Lemma 1, we recursively bound $\mathbb{E}[\|\lambda_K - \lambda^*\|^2]$ by

$$
\mathbb{E}[\|\lambda_K - \lambda^*\|^2] \leq 2Re^2
$$

for all $K \leq N$, which completes the proof. ∎

EC.2.4 Proof of Theorem 4

We set $\lambda_0 = 0$ throughout our analysis. We first consider the objective optimality gap $F(\bar{x}_N) - F(x^*)$. We denote by

$$
\hat{\pi}_2 \in \partial f_2(\bar{x}_N),\ \hat{\pi}_1 = \nabla f_1\Big(f_2(\bar{x}_N)\Big),\ \hat{v}_2 \in \partial g_2(\bar{x}_N),\ \text{and}\ \hat{v}_1 = \nabla g_1(g_2(\bar{x}_N)).
$$

Setting the feasible point as $(x^*, 0, \hat{\pi}_{1:2}, \hat{v}_{1:2})$ and following Theorem 2 with $M_\lambda = 0$ for $\lambda = 0$, $\alpha_t = 2\sqrt{N}$, and $\eta_t = \frac{15C_{g_1}^2C_{g_2}^2\sqrt{N}}{2}$, we obtain that

$$
\begin{aligned}
\mathbb{E}[F(\bar{x}_N)] - F(x^*) &\leq \frac{1}{N}\sum_{t=1}^{N}\mathbb{E}\Big(\mathcal{L}(x_t, 0, \hat{\pi}_{1:2}, \hat{v}_{1:2}) - \mathcal{L}(x^*, \lambda_t, \pi_{1:2,t}, v_{1:2,t})\Big) \\
&\leq \frac{1}{\sqrt{N}}\left(\frac{15C_{g_1}^2C_{g_2}^2}{4}\|x_0 - x^*\|^2 + \frac{7C_{f_1}^2C_{f_2}^2}{C_{g_1}^2C_{g_2^2}} + \frac{D_X^2C_{g_1}^2C_{g_2}^2 + \sigma_H^2}{2} + 2L_{f_1}\sigma_{f_2}^2 + 3C_{f_1}\sigma_{f_2}\right) \\
&\quad + \frac{1}{N}\sum_{t=1}^{N}\frac{1}{3\sqrt{N}}\Big(\mathbb{E}[\|\lambda_t\|^2] + \mathbb{E}[\|\lambda_{t-1}\|^2]\Big).
\end{aligned}
$$

We then derive the convergence rate of the optimality gap by the bound of $\|\lambda_t\|^2$ in Proposition 2 that

$$\mathbb{E}[\|\lambda_t\|^2] \le 2\|\lambda^*\|^2 + 2\mathbb{E}[\|\lambda_t - \lambda^*\|^2] \le 2\|\lambda^*\|^2 + 2Re^2, \text{ for all } t = 1, 2, \cdots, N.$$

Second, we consider the feasibility residual $G(\bar{x}_N)_+$. By adopting the feasible point $(x^*, \tilde{\lambda}, \hat{\pi}_{1:2}, \hat{v}_{1:2})$ where $\tilde{\lambda} = (\lambda^* + 1)\frac{G(\bar{x}_N)_+}{\|G(\bar{x}_N)_+\|_2}$ and following Theorem 2 with $M_{\tilde{\lambda}} = \|\tilde{\lambda}\|$, $\alpha_t = 2\sqrt{N}$, and $\eta_t = \frac{15C_{g_1}^2 C_{g_2}^2 \sqrt{N}}{2}$, we obtain

$$\begin{aligned}
\mathbb{E}[\|G(\bar{x}_N)_+\|_2] &\le \frac{1}{N}\sum_{t=1}^{N}\mathbb{E}\Big(\mathcal{L}(x_t, \tilde{\lambda}, \hat{\pi}_{1:2}, \hat{v}_{1:2}) - \mathcal{L}(x^*, \lambda_t, \pi_{1:2,t}, v_{1:2,t})\Big) \\
&\le \frac{1}{\sqrt{N}}\Big(\frac{16\|\tilde{\lambda}\|^2}{3} + 2L_{g_1}\sigma_{g_2}^2\|\tilde{\lambda}\| + 3\sqrt{K}C_{g_1}\sigma_{g_2}\|\tilde{\lambda}\| + \frac{15C_{g_1}^2 C_{g_2}^2}{4}\|x_0 - x^*\|^2\Big) \\
&\quad + \frac{1}{\sqrt{N}}\Big(\frac{7C_{f_1}^2 C_{f_2}^2}{C_{g_1}^2 C_{g_2}^2} + \frac{D_X^2 C_{g_1}^2 C_{g_2}^2 + \sigma_H^2}{2} + \|\tilde{\lambda}\|^2 + \sigma_H\|\tilde{\lambda}\| + 2L_{f_1}\sigma_{f_2}^2 + 3C_{f_1}\sigma_{f_2}\Big) \\
&\quad + \frac{1}{N}\sum_{t=1}^{N}\frac{1}{3\sqrt{N}}\Big(\mathbb{E}[\|\tilde{\lambda} - \lambda_t\|^2] + \mathbb{E}[\|\lambda_{t-1}\|^2]\Big).
\end{aligned}$$

We obtain the convergence rate of feasibility residual $\|G(x)_+\|_2$ by using the facts that

$$\|\lambda_t\|^2 \le 2\|\lambda^*\|^2 + 2\|\lambda_t - \lambda^*\|^2 \text{ and } \|\tilde{\lambda} - \lambda_t\|^2 \le 3\|\tilde{\lambda}\|^2 + 3\|\lambda^*\|^2 + 3\|\lambda_t - \lambda^*\|^2,$$

and applying $\mathbb{E}[\|\lambda_t - \lambda^*\|^2] \le 2Re^2$ provided by Proposition 2. This completes the proof. ∎

References

Lan G (2020) *First-order and Stochastic Optimization Methods for Machine Learning* (Springer Nature).

Zhang Z, Lan G (2024) Optimal methods for convex nested stochastic composite optimization. *Mathematical Programming* 1–48.