Lemma 3. *For AsySQN-SVRG, let $u \in K(t)$, we have that*

$$\begin{aligned}&\mathbb{E}\left\|\widehat{v}_u^\ell\right\|^2 \qquad (25)\\ &\le \frac{16L^2}{\mu}\mathbb{E}(f(w_t^s)-f(w^*)) + 4L^2\gamma^2\sigma_2^2\eta_1\sum_{v\in\{t,\dots,u\}}\mathbb{E}\left\|\widehat{v}_v^{\psi(v)}\right\|^2\\ &+\frac{8L^2}{\mu}\mathbb{E}(f(w^s)-f(w^*)) + 2\tau_1L^2\gamma^2\sigma_2^2\mathbb{E}\sum_{u'\in D(u)}\left\|\widehat{v}_{u'}^{\psi(u')}\right\|^2\end{aligned}$$

Define $v_u^\ell = \nabla_\ell f_{\mathcal{I}_u}(w_u^s) - \nabla_\ell f_{\mathcal{I}_u}(w^s) + \nabla_\ell f(w^s)$. We have that $\mathbb{E}\left\|\widehat{v}_u^\ell\right\|^2 = \mathbb{E}\left\|\widehat{v}_u^\ell - v_u^\ell + v_u^\ell\right\|^2 \le 2\mathbb{E}\left\|\widehat{v}_u^\ell - v_u^\ell\right\|^2 + 2\mathbb{E}\left\|v_u^\ell\right\|^2$. As for term $\mathbb{E}\left\|v_u^\ell\right\|^2$, we have that

$$\begin{aligned}&\mathbb{E}\left\|v_u^\ell\right\|^2 = \mathbb{E}\left\|\nabla_\ell f_{\mathcal{I}_u}(w_u^s) - \nabla_\ell f_{\mathcal{I}_u}(w^s) + \nabla_\ell f(w^s)\right\|^2\\ \le&2\mathbb{E}\left\|\nabla_\ell f_{\mathcal{I}_u}(w^s) - \nabla_\ell f_{\mathcal{I}_u}(w^*) - \nabla_\ell f(w^s) + \nabla_\ell f(w^*)\right\|^2\\ &+2\mathbb{E}\left\|\nabla_\ell f_{\mathcal{I}_u}(w_u^s) - \nabla_\ell f_{\mathcal{I}_u}(w^*)\right\|^2\\ \overset{(i)}{\le}&2\mathbb{E}\left\|\nabla_\ell f_{\mathcal{I}_u}(w_u^s) - \nabla_\ell f_{\mathcal{I}_u}(w^*)\right\|^2\\ &+2\mathbb{E}\left\|\nabla_\ell f_{\mathcal{I}_u}(w^s) - \nabla_\ell f_{\mathcal{I}_u}(w^*)\right\|^2\\ \overset{(ii)}{\le}&2L^2\mathbb{E}\left\|w_u^s - w^*\right\|^2 + 2L^2\mathbb{E}\left\|w^s - w^*\right\|^2\\ \le&4L^2\mathbb{E}\|w_u^s - w_t^s\|^2 + 4L^2\mathbb{E}\|w_t^s - w^*\|^2 + 2L^2\mathbb{E}\left\|w^s - w^*\right\|^2\\ \overset{(iii)}{=}&4L^2\gamma^2\mathbb{E}\|\sum_{v\in\{t,\dots,u\}}H_{\xi(v)}\widehat{v}_v^{\psi(v)}\|^2\\ &+4L^2\mathbb{E}\|w_t^s - w^*\|^2 + 2L^2\mathbb{E}\left\|w^s - w^*\right\|^2\\ \overset{(iv)}{\le}&\frac{8L^2}{\mu}\mathbb{E}(f(w_t^s)-f(w^*)) + \frac{4L^2}{\mu}\mathbb{E}(f(w^s)-f(w^*))\\ &+4L^2\gamma^2\sigma_2^2\eta_1\sum_{v\in\{t,\dots,u\}}\mathbb{E}\|\widehat{v}_v^{\psi(v)}\|^2 \qquad (26)\end{aligned}$$

where the (i) uses $\mathbb{E}\|x-\mathbb{E}x\|^2 \le \mathbb{E}\|x\|^2$, (ii) follows Eq. (22), (iii) uses Definition 2, and (iv) uses Assumption 1. Using $\|\sum_{i=1}^n a_i\|^2 \le n\sum_{i=1}^n \mathbb{E}\|a_i\|^2$, Definition 2, Assumption 4, and then we can bound $\mathbb{E}\left\|\widehat{v}_u^\ell - v_u^\ell\right\|^2$ as follows.

$$\mathbb{E}\left\|\widehat{v}_u^\ell - v_u^\ell\right\|^2 \le \tau_1L^2\sigma_2^2\gamma^2\mathbb{E}\sum_{u'\in D(u)}\left\|\widehat{v}_{u'}^{\psi(u')}\right\|^2, \qquad (27)$$

Combining $\mathbb{E}\left\|\widehat{v}_u^\ell\right\|^2 \le 2\mathbb{E}\left\|\widehat{v}_u^\ell - v_u^\ell\right\|^2 + 2\mathbb{E}\left\|v_u^\ell\right\|^2$ with Eqs. (26) to (27), we have the final result as in Eq. 25. Similar to the proof of (20), for $u \in K(t)$ at s-th outer loop, we have

$$\begin{aligned}&\mathbb{E}f(w_{u+1}^s) \qquad (28)\\ &\le \mathbb{E}(f(w_u^s) + \langle\nabla f(w_u^s), w_{u+1}^s - w_u^s\rangle + \frac{L_\ell}{2}\|w_{u+1}^s - w_u^s\|^2)\\ &\le \mathbb{E}f(w_u^s) - \gamma\mathbb{E}\langle\nabla f(w_u^s), H^\ell\nabla_\ell f_{\mathcal{I}_u}(\widehat{w}_u^s) - H^\ell\nabla_\ell f_{\mathcal{I}_u}(w_u^s)\\ &+H^\ell\nabla_\ell f_{\mathcal{I}_u}(w_u^s)\rangle + \frac{L_\ell\gamma^2\|H^\ell\|^2}{2}\mathbb{E}\|\widehat{v}_u^\ell\|^2\\ &\le\mathbb{E}f(w_u^s) - \frac{\gamma\sigma_1}{2}\mathbb{E}\|\nabla_\ell f(w_u^s)\|^2 + \frac{\tau_1\sigma_2^3L^2\gamma^3}{2}\sum_{u'\in D(u)}\mathbb{E}\|\widehat{v}_{u'}^{\psi(u')}\|^2\\ &+\frac{L_\ell\gamma^2\sigma_2^2}{2}\mathbb{E}\|\widehat{v}_u^\ell\|^2 \qquad (29)\end{aligned}$$

Summing Eq. (28) over all $u \in K(t)$, and using Lemma 2, Assumption 5, and Lemma 3, we have

$$\begin{aligned}&\mathbb{E}f(w_{t+|K(t)|}^s) - \mathbb{E}f(w_t^s) \le -\frac{\gamma\mu\sigma_1}{2}\mathbb{E}(f(w_t^s)-f(w^*)) \qquad (30)\\ &+\frac{\tau_1\sigma_2^3L^2\gamma^3}{2}\sum_{u\in K(t)}\sum_{u'\in D(u)}\mathbb{E}\|\widehat{v}_{u'}^{\psi(u')}\|^2 + C\\ &\cdot\sum_{u\in K(t)}\Bigg(\frac{16L^2}{\mu}\mathbb{E}(f(w_t^s)-f(w^*)) + \frac{8L^2}{\mu}\mathbb{E}(f(w^s)-f(w^*))\\ &+4\alpha\sum_{v\in\{t,\dots,u\}}\mathbb{E}\left\|\widehat{v}_v^{\psi(v)}\right\|^2 + 2\tau_1L^2\gamma^2\sigma_2^2\mathbb{E}\sum_{u'\in D(u)}\left\|\widehat{v}_{u'}^{\psi(u')}\right\|^2\Bigg)\end{aligned}$$

Let $e_t^s = \mathbb{E}(f(w_t^s)-f(w^*))$ and $e^s = \mathbb{E}(f(w^s)-f(w^*))$, we have

$$\begin{aligned}e_{t+|K(t)|}^s \le& \left(1-\frac{\gamma\mu\sigma_1}{2}+\frac{16L^2\eta_1C}{\mu}\right)e_t^s + \frac{8L^2\eta_1C}{\mu}e^s \qquad (31)\\ &+\gamma^3\left(\left(\frac{\sigma_2}{2}+\frac{2C}{\gamma}\right)\tau_1^2 + 4\frac{C}{\gamma}\eta_1^2\right)\eta_1L^2\sigma_2^2\frac{9G}{b}\end{aligned}$$

We carefully choose γ such that $1 > \frac{\gamma\mu}{2} - \frac{16L^2\eta_1C}{\mu} \overset{\text{def}}{=} \rho > 0$. Assume that $\cup_{\kappa\in P(t)} = \{0,1,\dots,t\}$, applying (31), we have that

$$\begin{aligned}e_t^s \le& (1-\rho)^{v(t)}e^s + \left(\frac{8L^2\eta_1C}{\mu}e^s\right)\sum_{i=0}^{v(t)}(1-\rho)^i \qquad (32)\\ &+(\gamma^3\left(\left(\frac{\sigma_2}{2}+\frac{2C}{\gamma}\right)\tau_1^2 + 4\frac{C}{\gamma}\eta_1^2\right)\eta_1qL^2\sigma_2^2\frac{9G}{b})\sum_{i=0}^{v(t)}(1-\rho)^i\\ \le& (1-\rho)^{v(t)}e^s + \left(\frac{8L^2\eta_1C}{\mu}e^s\right)\frac{1}{\rho}\\ &+(\gamma^3\left(\left(\frac{\sigma_2}{2}+\frac{2C}{\gamma}\right)\tau_1^2 + 4\frac{C}{\gamma}\eta_1^2\right)\eta_1qL^2\sigma_2^2\frac{9G}{b})\frac{1}{\rho}\\ =&\left((1-\rho)^{v(t)} + \frac{8L^2\eta_1C}{\rho\mu}\right)e^s\\ &+\gamma^3\left(\left(\frac{\sigma_2}{2}+\frac{2C}{\gamma}\right)\tau_1^2 + 4\frac{C}{\gamma}\eta_1^2\right)\frac{9\eta_1L^2\sigma_2^2G}{b\rho}\end{aligned}$$

Thus, to achieve $\mathbb{E}f(w_S) - f(w^*) \le \epsilon$, we can carefully choose γ such that 1) $\gamma^3\left(\left(\frac{\sigma_2}{2}+\frac{2C}{\gamma}\right)\tau_1^2 + 4\frac{C}{\gamma}\eta_1^2\right)\frac{9\eta_1L^2\sigma_2^2G}{b\rho} \le \frac{\epsilon}{8}$ and 2) $\frac{8L^2\eta_1C}{\rho\mu} \le 0.5$; and let $(1-\rho)^{v(t)} \le 0.25$, i.e., $v(t) \ge \frac{\log 0.25}{\log(1-\rho)}$, we have that

$$e^{s+1} \le 0.75e^s + \frac{\epsilon}{8} \qquad (33)$$

Recursively apply (33), we have that

$$e^S \le (0.75)^S e^0 + \frac{\epsilon}{2}$$

Finally, we use that the outer loop number S should satisfy the condition of $S \ge \frac{\log\frac{2e^0}{\epsilon}}{\log\frac{4}{3}}$ and can obtain the desired result in Theorem 2. This completes the proof.