# Appendix for Paper “Asynchronous Doubly Stochastic Group Regularized Learning”

**Bin Gu** BIN.GU@PITT.EDU
**Zhouyuan Huo** ZHOUYUAN.HUO@PITT.EDU
**Heng Huang** HENG.HUANG@PITT.EDU
*Department of Electrical & Computer Engineering, University of Pittsburgh, USA*

## 1. Convergence Analysis

In this section, we follow the analysis of (Liu and Wright, 2015) and prove the convergence rate of AsyDSPG+ (Theorems 5 and 6). Specifically, AsyDSPG+ achieves a linear convergence rate when the function $f$ is with the optimal strong convexity property, and a sublinear rate when $f$ is with the general convexity (Theorems 5). In addition, AsyDSPG+ alsos achieves a sublinear rate when $f$ is with the non-convexity (Theorems 6).

Before providing the theoretical analysis, we give the definitions of $\widehat{x}^s_{t,t'+1}$, $\overline{x}^s_{t+1}$, $\widetilde{\nabla} F(x^s_t)$ and the explanation of $x^s_t$ used in the analysis as follows.

1. $\widehat{x}^s_{t,t'}$**:** Assume the indices in $K(t)$ are sorted in the increasing order, we use $K(t)_{t'}$ to denote the $t'$-th index in $K(t)$. For $t' = 0, 1, \cdots, |K(t)|$, we define

$$\widehat{x}^s_{t,t'} = \widehat{x}^s_t + \sum_{t''=1}^{t'} \left( B^s_{K(t)_{t''}} \Delta^s_{K(t)_{t''}} \right) = \widehat{x}^s_t + \sum_{t''=1}^{t'} \left( x_{K(t)_{t''}+1} - x_{K(t)_{t''}} ) \right) \tag{20}$$

Thus, we have that

$$\widehat{x}^s_t = \widehat{x}^s_{t,0} \tag{21}$$

$$x^s_t = \widehat{x}^s_{t,|K(t)|} \tag{22}$$

$$x^s_t - \widehat{x}^s_t = \sum_{t''=0}^{|K(t)|-1} \left( B^{s+1}_{K(t)_{t''}} \Delta^{s+1}_{K(t)_{t''}} \right) = \sum_{t'=0}^{|K(t)|-1} \left( \widehat{x}^s_{t,t'+1} - \widehat{x}^s_{t,t'} \right) \tag{23}$$

$$\nabla f(x^s_t) - \nabla f(\widehat{x}^s_t) = \sum_{t'=0}^{|K(t)|-1} \left( \nabla f(\widehat{x}^s_{t,t'+1}) - \nabla f(\widehat{x}_{t,t'}) \right) \tag{24}$$

2. $\overline{x}^s_{t+1}$**:** $\overline{x}^s_{t+1}$ is defined as:

$$\overline{x}^s_{t+1} \stackrel{\text{def}}{=} \mathcal{P}_{\frac{\gamma}{L_{\max}} g} \left( x^s_t - \frac{\gamma}{L_{\max}} \widehat{v}^s_t \right) \tag{25}$$

Based on (25), it is easy to verify that $(\overline{x}^s_{t+1})_{\mathcal{G}_{j(t)}} = (x^{s+1}_{t+1})_{\mathcal{G}_{j(t)}}$. Thus, we have $\mathbb{E}_{j(t)}(x^s_{t+1} - x^s_t) = \frac{1}{k}\left(\overline{x}^s_{t+1} - x^s_t\right)$. It means that $\overline{x}^s_{t+1} - x^s_t$ captures the expectation of $x^s_{t+1} - x^s_t$.

3. $\widetilde{\nabla} F(x^s_t)$**:** If the function $f$ is a non-convex function, AsyDSPG+ may not converge to the global optimum point. Thus, the closeness to the optimal solution (i.e., $F(x) - F^*$ and $\|x - \mathcal{P}_{S(x)}\|$) cannot be used for the convergence analysis. To analyze the convergence rate of AsyDSPG+ in the non-convex setting, same with (Razaviyayn et al., 2014; Yu and Tao, 2016), we use $\mathbb{E}\widetilde{\nabla} F(x^s_t)$ as defined in (26).

$$\mathbb{E}\widetilde{\nabla} F(x^s_t) \stackrel{\text{def}}{=} \frac{L_{\max}}{\gamma}\mathbb{E}\left(x^s_t - \overline{x}^s_{t+1}\right) \tag{26}$$

It is easy to verify that $\mathbb{E}\widetilde{\nabla} F(x^s_t)$ will equal to $\mathbf{0}$ when AsyDSPG+ approaches to a stationary point. Note that when $g(x) = 0$, we have that $\mathbb{E}\widetilde{\nabla} F(x^s_t) = \nabla f(\widehat{x}^s_t)$.

4. $x^s_t$**:** As mentioned previously, AsyDSPG+ does not use any locks in the reading and writing. Thus, in the line 10 of Algorithm 1, $x^s_t$ (left side of '$\leftarrow$') updated in the shared memory may be inconsistent with the ideal one (right side of '$\leftarrow$') computed by the proximal operator. In the analysis, we use $x^s_t$ to denote the ideal one computed by the proximal operator. Same as mentioned in (Mania et al., 2015), there might not be an actual time the ideal ones exist in the shared memory, except the first and last iterates for each outer loop. It is noted that, $x^s_0$ and $x^s_m$ are exactly what is stored in shared memory. Thus, we only consider the ideal $x^s_t$ in the analysis.

Then, we give two inequalities in Lemma 1 and 2 respectively. Based on Lemma 1 and 2, we prove that $\mathbb{E}\|x^s_{t-1} - \overline{x}^s_t\|^2 \leq \rho\mathbb{E}\|x^s_t - \overline{x}^s_{t+1}\|^2$ (Lemma 3), where $\rho > 1$ is a user defined parameter. Then, we prove the monotonicity of the expectation of the objectives $\mathbb{E}F(x^s_{t+1}) \leq \mathbb{E}F(x^s_t)$ (Lemma 4). Note that the analyses only consider the case $|\mathcal{B}| = 1$ without loss of generality. The case of $|\mathcal{B}| > 1$ can be proved similarly.

**Lemma 1** *For $\|\nabla f(x^s_t) - \nabla f(\widehat{x}^s_t)\|$ in each iteration of AsyDSPG+, we have its upper bound as*

$$\|\nabla f(x^s_t) - \nabla f(\widehat{x}^s_t)\| \leq L_{res}\sum_{t' \in K(t)}\|\Delta^s_{t'}\| \tag{27}$$

**Proof** Based on (24), we have that

$$\|\nabla f(x^s_t) - \nabla f(\widehat{x}^s_t)\| = \left\|\sum_{t'=0}^{|K(t)|-1}\nabla f(\widehat{x}^s_{t,t'+1}) - \nabla f(\widehat{x}^s_{t,t'})\right\| \tag{28}$$

$$\leq \sum_{t'=0}^{|K(t)|-1}\left\|\nabla f(\widehat{x}^s_{t,t'+1}) - \nabla f(\widehat{x}^s_{t,t'})\right\| \leq L_{res}\sum_{t'=0}^{|K(t)|-1}\left\|\widehat{x}^s_{t,t'+1} - \widehat{x}^s_{t,t'}\right\|$$

$$= L_{res}\sum_{t'=0}^{|K(t)|-1}\left\|B^s_{K(t)_{t'}}\Delta^s_{K(t)_{t'}}\right\| \leq L_{res}\sum_{t'=0}^{|K(t)|-1}\left\|B^s_{K(t)_{t'}}\right\|\left\|\Delta^s_{K(t)_{t'}}\right\| \leq L_{res}\sum_{t' \in K(t)}\|\Delta^s_{t'}\|$$

This completes the proof. ■

**Lemma 2** *In each iteration of AsyDSPG+, $\forall x$, we have the following inequality.*

$$\left\langle (\widehat{v}_t^s)_{\mathcal{G}_{j(t)}} + \frac{L_{\max}}{\gamma}\Delta_t^s, (x_{t+1}^s - x)_{\mathcal{G}_{j(t)}} \right\rangle + g_{\mathcal{G}_{j(t)}}\left((x_{t+1}^s)_{\mathcal{G}_{j(t)}}\right) - g_{\mathcal{G}_{j(t)}}\left((x)_{\mathcal{G}_{j(t)}}\right) \leq 0 \tag{29}$$

**Proof** The problem solved in lines 8 of Algorithm 1 is as follows

$$\begin{aligned} x_{t+1}^s = \arg\min_x \quad & \left\langle (\widehat{v}_t^s)_{\mathcal{G}_{j(t)}}, (x - x_t^s)_{\mathcal{G}_{j(t)}} \right\rangle + \frac{L_{\max}}{2\gamma}\left\|(x - x_t^s)_{\mathcal{G}_{j(t)}}\right\|^2 \\ & + g_{\mathcal{G}_{j(t)}}\left((x)_{\mathcal{G}_{j(t)}}\right) \\ s.t. \quad & x_{\backslash \mathcal{G}_{j(t)}} = (x_t^s)_{\backslash \mathcal{G}_{j(t)}} \end{aligned} \tag{30}$$

If $x_{t+1}^s$ is the solution of (30), the solution of optimization problem (31) is also $x_{t+1}^s$ according to the subdifferential version of Karush-Kuhn-Tucker (KKT) conditions (Ruszczyński., 2006).

$$\begin{aligned} P(x) = \min_x \quad & \left\langle (\widehat{v}_t^s)_{\mathcal{G}_{j(t)}} + \frac{L_{\max}}{\gamma}\left(x_{t+1}^s - x_t^s\right)_{\mathcal{G}_{j(t)}}, (x - x_t^s)_{\mathcal{G}_{j(t)}} \right\rangle + g_{\mathcal{G}_{j(t)}}\left((x)_{\mathcal{G}_{j(t)}}\right) \\ s.t. \quad & x_{\backslash \mathcal{G}_{j(t)}} = (x_t^s)_{\backslash \mathcal{G}_{j(t)}} \end{aligned} \tag{31}$$

Thus, we have that $P(x) \geq P(x_{t+1}^s)$, $\forall x$, which leads to (29). This completes the proof. ■

**Lemma 3** *Let $\rho$ be a constant that satisfies $\rho > 1$, and define the quantities $\theta_1 = \frac{\rho^{\frac{1}{2}} - \rho^{\frac{\tau+1}{2}}}{1-\rho^{\frac{1}{2}}}$ and $\theta_2 = \frac{\rho^{\frac{1}{2}} - \rho^{\frac{m}{2}}}{1-\rho^{\frac{1}{2}}}$. Suppose the nonnegative steplength parameter $\gamma > 0$ satisfies $\gamma \leq \min\left\{\frac{k^{1/2}(1-\rho^{-1})-4}{4(\Lambda_{res}(1+\theta_1)+\Lambda_{nor}(1+\theta_2))}, \frac{k^{1/2}}{\frac{1}{2}k^{1/2}+2\Lambda_{nor}\theta_2+\Lambda_{res}\theta_1}\right\}$, under Assumptions 1, 2, 3 and 4, we have*

$$\mathbb{E}\|x_{t-1}^s - \overline{x}_t^s\|^2 \leq \rho\mathbb{E}\|x_t^s - \overline{x}_{t+1}^s\|^2 \tag{32}$$

**Proof** According to (A.8) in (Liu and Wright, 2015), we have

$$\|x_{t-1}^s - \overline{x}_t^s\|^2 - \|x_t^s - \overline{x}_{t+1}^s\|^2 \leq 2\|x_{t-1}^s - \overline{x}_t^s\|\|x_t^s - \overline{x}_{t+1}^s - x_{t-1}^s + \overline{x}_t^s\| \tag{33}$$

The second part in the right half side of (33) is bound as follows if $\mathcal{B} = \{i_t\}$ and $J(t) = \{j(t)\}$.

$$\begin{aligned} & \|x_t^s - \overline{x}_{t+1}^s - x_{t-1}^s + \overline{x}_t^s\| \qquad (34) \\ = \ & \left\|x_t^s - \mathcal{P}_{\frac{\gamma}{L_{\max}}g}\left(x_t^s - \frac{\gamma}{L_{\max}}\widehat{v}_t^s\right) - x_{t-1}^s + \mathcal{P}_{\frac{\gamma}{L_{\max}}g}\left(x_{t-1}^s - \frac{\gamma}{L_{\max}}\widehat{v}_{t-1}^s\right)\right\| \\ \leq \ & \|x_t^s - x_{t-1}^s\| + \left\|\mathcal{P}_{\frac{\gamma}{L_{\max}}g}\left(x_t^s - \frac{\gamma}{L_{\max}}\widehat{v}_t^s\right) - \mathcal{P}_{\frac{\gamma}{L_{\max}}g}\left(x_{t-1}^s - \frac{\gamma}{L_{\max}}\widehat{v}_{t-1}^s\right)\right\| \end{aligned}$$

$$
\begin{aligned}
&\leq 2\|x_t^s - x_{t-1}^s\| + \frac{\gamma}{L_{\max}} \left\|\widehat{v}_t^s - \widehat{v}_{t-1}^s\right\| \\
&= 2\|x_t^s - x_{t-1}^s\| + \\
&\quad \frac{\gamma}{L_{\max}} \left\|\nabla f_{i_t}(\widehat{x}_t^s) - \nabla f_{i_t}(\widetilde{x}^{s-1}) + \nabla f(\widetilde{x}^{s-1}) - \nabla f_{i_{t-1}}(\widehat{x}_{t-1}^s) + \nabla f_{i_{t-1}}(\widetilde{x}^{s-1}) - \nabla f(\widetilde{x}^{s-1})\right\| \\
&= 2\|x_t^s - x_{t-1}^s\| + \frac{\gamma}{L_{\max}} \left\|\nabla f_{i_t}(\widehat{x}_t^s) - \nabla f_{i_t}(\widetilde{x}^{s-1}) - \nabla f_{i_{t-1}}(\widehat{x}_{t-1}^s) + \nabla f_{i_{t-1}}(\widetilde{x}^{s-1})\right\| \\
&\leq 2\|x_t^s - x_{t-1}^s\| + \frac{\gamma}{L_{\max}} \left\|\nabla f_{i_t}(\widehat{x}_t^s) - \nabla f_{i_t}(x_t^s) + \nabla f_{i_t}(x_t^s) - \nabla f_{i_t}(\widetilde{x}^{s-1})\right\| \\
&\quad + \frac{\gamma}{L_{\max}} \left\|\nabla f_{i_{t-1}}(\widehat{x}_{t-1}^s) - \nabla f_{i_{t-1}}(x_{t-1}^s) + \nabla f_{i_{t-1}}(x_{t-1}^s) - \nabla f_{i_{t-1}}(\widetilde{x}^{s-1})\right\| \\
&\leq 2\|x_t^s - x_{t-1}^s\| + \frac{\gamma}{L_{\max}} \|\nabla f_{i_t}(\widehat{x}_t^s) - \nabla f_{i_t}(x_t^s)\| + \frac{\gamma}{L_{\max}} \left\|\nabla f_{i_t}(x_t^s) - \nabla f_{i_t}(\widetilde{x}^{s-1})\right\| \\
&\quad + \frac{\gamma}{L_{\max}} \left\|\nabla f_{i_{t-1}}(\widehat{x}_{t-1}^s) - \nabla f_{i_{t-1}}(x_{t-1}^s)\right\| + \frac{\gamma}{L_{\max}} \left\|\nabla f_{i_{t-1}}(x_{t-1}^s) - \nabla f_{i_{t-1}}(\widetilde{x}^{s-1})\right\| \\
&\leq 2\|x_t^s - x_{t-1}^s\| + \gamma\Lambda_{res} \left( \sum_{t' \in K(t-1)} \|\Delta_{t'}^s\| + \sum_{t' \in K(t)} \|\Delta_{t'}^s\| \right) \\
&\quad + \gamma\Lambda_{nor} \left( \|x_t^s - \widetilde{x}^s\| + \|x_{t-1}^s - \widetilde{x}^s\| \right) \\
&\leq 2\|x_t^s - x_{t-1}^s\| + 2\gamma \left( \Lambda_{res} \sum_{t'=t-1-\tau}^{t-1} \|\Delta_{t'}^s\| + \Lambda_{nor} \sum_{t'=0}^{t-1} \|\Delta_{t'}^s\| \right)
\end{aligned}
$$

where the first inequality use the nonexpansive property of $\mathcal{P}_{\frac{\gamma}{L_{\max}} g}$, the fifth inequality use A.7 of (Liu and Wright, 2015), the sixth inequality comes from $\|x_t^s - \widetilde{x}^s\| = \|\sum_{t'=0}^{t-1} \Delta_{t'}^s\| \leq \sum_{t'=0}^{t-1} \|\Delta_{t'}^s\|$.

If $t = 1$, we have that $K(0) = \emptyset$ and $K(1) \subseteq \{0\}$. Thus, according to (34), we have

$$
\|x_1^s - \overline{x}_2^s - x_0^s + \overline{x}_1^s\| \leq 2\|x_1^s - x_0^s\| + 2\gamma \left(\Lambda_{res} + \Lambda_{nor}\right) \|\Delta_0^s\| \tag{35}
$$

Substituting (35) into (33), and takeing expectations, we have

$$
\begin{aligned}
&\mathbb{E}\|x_0^s - \overline{x}_1^s\|^2 - \mathbb{E}\|x_1^s - \overline{x}_2^s\|^2 \leq 2\mathbb{E}\left(\|x_0^s - \overline{x}_1^s\| \|x_1^s - \overline{x}_2^s - x_0^s + \overline{x}_1^s\|\right) \qquad (36) \\
\leq\ & 4\mathbb{E}\left(\|x_0^s - \overline{x}_1^s\| \|x_1^s - x_0^s\|\right) + 4\gamma \left(\Lambda_{res} + \Lambda_{nor}\right) \mathbb{E}\left(\|x_0^s - \overline{x}_1^s\| \|\Delta_0^s\|\right) \\
\leq\ & 4k^{-\frac{1}{2}} \mathbb{E}\left(\|x_0^s - \overline{x}_1^s\|^2\right) + 4\gamma \left(\Lambda_{res} + \Lambda_{nor}\right) \mathbb{E}\left(\|x_0^s - \overline{x}_1^s\| \|\Delta_0^s\|\right)
\end{aligned}
$$

where the last inequality uses A.13 in (Liu and Wright, 2015). Further, we have the upper bound of $\mathbb{E}\left(\|x_t^s - \overline{x}_{t+1}^s\| \|\Delta_t^s\|\right)$ as

$$
\begin{aligned}
&\mathbb{E}\left(\|x_t^s - \overline{x}_{t+1}^s\| \|\Delta_t^s\|\right) \leq \frac{1}{2}\mathbb{E}\left(k^{-\frac{1}{2}} \|x_t^s - \overline{x}_{t+1}^s\|^2 + k^{\frac{1}{2}} \|\Delta_t^s\|^2\right) \qquad (37) \\
=\ & \frac{1}{2}\mathbb{E}\left(k^{-\frac{1}{2}} \|x_t^s - \overline{x}_{t+1}^s\|^2 + k^{\frac{1}{2}} \mathbb{E}_{j(t)} \|\Delta_t^s\|^2\right) = \frac{1}{2}\mathbb{E}\left(k^{-\frac{1}{2}} \|x_t^s - \overline{x}_{t+1}^s\|^2 + k^{-\frac{1}{2}} \mathbb{E}\|x_t^s - \overline{x}_{t+1}^s\|^2\right) \\
=\ & k^{-\frac{1}{2}} \mathbb{E}\|x_t^s - \overline{x}_{t+1}^s\|^2
\end{aligned}
$$

Substituting (37) into (36), we have

$$
\mathbb{E}\|x_0^s - \overline{x}_1^s\|^2 - \mathbb{E}\|x_1^s - \overline{x}_2^s\|^2 \leq k^{-\frac{1}{2}} \left(4 + 4\gamma \left(\Lambda_{res} + \Lambda_{nor}\right)\right) \mathbb{E}\left(\|x_0^s - \overline{x}_1^s\|^2\right) \tag{38}
$$

which implies that

$$\mathbb{E}\|x_0^s - \overline{x}_1^s\|^2 \leq \left(1 - \frac{4 + 4\gamma\left(\Lambda_{res} + \Lambda_{nor}\right)}{\sqrt{k}}\right)^{-1} \mathbb{E}\|x_1^s - \overline{x}_2^s\|^2 \leq \rho\mathbb{E}\|x_1^s - \overline{x}_2^s\|^2 \quad (39)$$

where the last inequality follows from . Thus, we have (32) for $t = 1$.

$$\rho^{-1} \leq 1 - \frac{4 + 4\gamma\left(\Lambda_{res} + \Lambda_{nor}\right)}{\sqrt{k}} \Leftrightarrow \gamma \leq \frac{k^{1/2}(1 - \rho^{-1}) - 4}{4\left(\Lambda_{res} + \Lambda_{nor}\right)} \quad (40)$$

Next, we consider the cases for $t > 1$. For $t - 1 - \tau \leq t' \leq t - 1$ and any $\beta > 0$, we have

$$\begin{aligned}
& \mathbb{E}\left(\|x_t^s - \overline{x}_{t+1}^s\|\|\Delta_{t'}^s\|\right) \leq \frac{1}{2}\mathbb{E}\left(k^{-\frac{1}{2}}\beta^{-1}\|x_t^s - \overline{x}_{t+1}^s\|^2 + k^{\frac{1}{2}}\beta\|\Delta_{t'}^s\|^2\right) \quad (41)\\
= \quad & \frac{1}{2}\mathbb{E}\left(k^{-\frac{1}{2}}\beta^{-1}\|x_t^s - \overline{x}_{t+1}^s\|^2 + k^{\frac{1}{2}}\beta\mathbb{E}_{j(t)}\|\Delta_{t'}^s\|^2\right)\\
= \quad & \frac{1}{2}\mathbb{E}\left(k^{-\frac{1}{2}}\beta^{-1}\|x_t^s - \overline{x}_{t+1}^s\|^2 + k^{-\frac{1}{2}}\beta\mathbb{E}\|x_{t'}^s - \overline{x}_{t'+1}^s\|^2\right)\\
\leq \quad & \frac{1}{2}\mathbb{E}\left(k^{-\frac{1}{2}}\beta^{-1}\|x_t^s - \overline{x}_{t+1}^s\|^2 + k^{-\frac{1}{2}}\rho^{t-t'}\beta\mathbb{E}\|x_t^s - \overline{x}_{t+1}^s\|^2\right)\\
\overset{\beta=\rho^{\frac{t'-t}{2}}}{\leq} \quad & k^{-\frac{1}{2}}\rho^{\frac{t-t'}{2}}\mathbb{E}\|x_t^s - \overline{x}_{t+1}^s\|^2
\end{aligned}$$

We assume that (32) holds $\forall t' < t$. By substituting (34) into (33) and taking expectation on both sides of (33), we can have

$$\begin{aligned}
& \mathbb{E}\left(\|x_{t-1}^s - \overline{x}_t^s\|^2 - \|x_t^s - \overline{x}_{t+1}^s\|^2\right) \quad (42)\\
\leq \quad & 2\mathbb{E}\left(\|x_{t-1}^s - \overline{x}_t^s\|\|x_t^s - \overline{x}_{t+1} - x_{t-1}^s + \overline{x}_t^s\|\right)\\
\leq \quad & 2\mathbb{E}\left(\|x_{t-1}^s - \overline{x}_t^s\|\left(2\|x_t^s - x_{t-1}^s\| + 2\gamma\left(\Lambda_{res}\sum_{t'=t-1-\tau}^{t-1}\|\Delta_{t'}^s\| + \Lambda_{nor}\sum_{t'=0}^{t-1}\|\Delta_{t'}^s\|\right)\right)\right)\\
= \quad & 4\mathbb{E}\left(\|x_{t-1}^s - \overline{x}_t^s\|\|x_{t-1}^s - \overline{x}_t^s\|\right) +\\
& 4\gamma\mathbb{E}\left(\Lambda_{res}\sum_{t'=t-1-\tau}^{t-1}\|x_{t-1}^s - \overline{x}_t^s\|\|\Delta_{t'}^s\| + \Lambda_{nor}\sum_{t'=0}^{t-1}\|x_{t-1}^s - \overline{x}_t^s\|\|\Delta_{t'}^s\|\right)\\
\leq \quad & 4k^{-1/2}\mathbb{E}\left(\|x_{t-1}^s - \overline{x}_t^s\|^2\right) +\\
& 4\gamma k^{-1/2}\mathbb{E}\left(\|x_{t-1}^s - \overline{x}_t^s\|^2\right)\cdot\left(\Lambda_{res}\sum_{t'=t-1-\tau}^{t-1}\rho^{\frac{t-1-t'}{2}} + \Lambda_{nor}\sum_{t'=0}^{t-1}\rho^{\frac{t-1-t'}{2}}\right)\\
= \quad & (4 + 4\gamma\left(\Lambda_{res} + \Lambda_{nor}\right))k^{-1/2}\mathbb{E}\left(\|x_{t-1}^s - \overline{x}_t^s\|^2\right) +\\
& 4\gamma k^{\frac{-1}{2}}\mathbb{E}\left(\|x_{t-1}^s - \overline{x}_t^s\|^2\right)\left(\Lambda_{res}\sum_{t'=1}^{\tau}\rho^{\frac{t'}{2}} + \Lambda_{nor}\sum_{t'=1}^{t-1}\rho^{\frac{t'}{2}}\right)\\
= \quad & k^{-1/2}\mathbb{E}\left(\|x_{t-1}^s - \overline{x}_t^s\|^2\right)\left(4 + 4\gamma\Lambda_{res}\left(1 + \frac{\rho^{\frac{1}{2}} - \rho^{\frac{\tau+1}{2}}}{1 - \rho^{\frac{1}{2}}}\right) + 4\gamma\Lambda_{nor}\left(1 + \frac{\rho^{\frac{1}{2}} - \rho^{\frac{m}{2}}}{1 - \rho^{\frac{1}{2}}}\right)\right)\\
= \quad & k^{-1/2}\mathbb{E}\left(\|x_{t-1}^s - \overline{x}_t^s\|^2\right)\cdot\left(4 + 4\gamma\left(\Lambda_{res}\left(1 + \theta_1\right) + \Lambda_{nor}\left(1 + \theta_2\right)\right)\right)
\end{aligned}$$

where the third inequality uses (41). Based on (42), we have that

$$
\begin{aligned}
& \mathbb{E}\left(\|x_{t-1}^s - \overline{x}_t^s\|^2\right) \qquad (43)\\
\leq\ & \left(1 - k^{-1/2}\left(4 + 4\gamma\left(\Lambda_{res}\left(1+\theta_1\right) + \Lambda_{nor}\left(1+\theta_2\right)\right)\right)\right)^{-1} \cdot \mathbb{E}\left(\|x_t^s - \overline{x}_{t+1}^s\|^2\right) \\
\leq\ & \rho\mathbb{E}\left(\|x_t^s - \overline{x}_{t+1}^s\|^2\right)
\end{aligned}
$$

where the last inequality follows from

$$
\begin{aligned}
\rho^{-1} \leq 1 - k^{-1/2}\left(4 + 4\gamma\left(\Lambda_{res}\left(1+\theta_1\right) + \Lambda_{nor}\left(1+\theta_2\right)\right)\right) \qquad (44)\\
\Leftrightarrow \gamma \leq \frac{k^{1/2}(1-\rho^{-1}) - 4}{4\left(\Lambda_{res}\left(1+\theta_1\right) + \Lambda_{nor}\left(1+\theta_2\right)\right)}
\end{aligned}
$$

This completes the proof. ■

**Lemma 4** *Let $\rho$ be a constant that satisfies $\rho > 1$, and define the quantities $\theta_1 = \frac{\rho^{\frac{1}{2}} - \rho^{\frac{\tau+1}{2}}}{1-\rho^{\frac{1}{2}}}$ and $\theta_2 = \frac{\rho^{\frac{1}{2}} - \rho^{\frac{m}{2}}}{1-\rho^{\frac{1}{2}}}$. Suppose the nonnegative steplength parameter $\gamma > 0$ satisfies $\gamma \leq \min\left\{\frac{k^{1/2}(1-\rho^{-1})-4}{4(\Lambda_{res}(1+\theta_1)+\Lambda_{nor}(1+\theta_2))}, \frac{k^{1/2}}{\frac{1}{2}k^{1/2}+2\Lambda_{nor}\theta_2+\Lambda_{res}\theta_1}\right\}$. Under Assumptions 1, 2, 3 and 4, the expectation of the objective function $\mathbb{E}F(x_t^s)$ is monotonically decreasing, i.e., $\mathbb{E}F(x_{t+1}^s) \leq \mathbb{E}F(x_t^s)$.*

**Proof** Take expectation $F(x_{t+1}^s)$ on $j(t)$, we have that

$$
\begin{aligned}
& \mathbb{E}_{j(t)}F(x_{t+1}^s) = \mathbb{E}_{j(t)}\left(f(x_t^s + \Delta_t^s) + g(x_{t+1}^s)\right) \qquad (45)\\
\leq\ & \mathbb{E}_{j(t)}\Bigg(f(x_t^s) + \left\langle \nabla_{\mathcal{G}_{j(t)}} f(x_t^s), (\Delta_t^s)_{\mathcal{G}_{j(t)}}\right\rangle + \frac{L_{\max}}{2}\left\|(\Delta_t^s)_{\mathcal{G}_{j(t)}}\right\|^2 \\
& + g_{\mathcal{G}_{j(t)}}\left((x_{t+1}^s)_{\mathcal{G}_{j(t)}}\right) + \sum_{j' \neq j(t)} g_{\mathcal{G}_{j'}}\left((x_{t+1}^s)_{\mathcal{G}_{j'}}\right)\Bigg) \\
=\ & f(x_t^s) + \frac{k-1}{k}g(x_t^s) + \mathbb{E}_{j(t)}\Bigg(\left\langle \nabla_{\mathcal{G}_{j(t)}} f(x_t^s), (\Delta_t^s)_{\mathcal{G}_{j(t)}}\right\rangle + \frac{L_{\max}}{2}\left\|(\Delta_t^s)_{\mathcal{G}_{j(t)}}\right\|^2 \\
& + g_{\mathcal{G}_{j(t)}}\left((x_{t+1}^s)_{\mathcal{G}_{j(t)}}\right)\Bigg) \\
=\ & F(x_t^s) + \mathbb{E}_{j(t)}\Bigg(\left\langle (\widehat{v}_t^s)_{\mathcal{G}_{j(t)}}, (\Delta_t^s)_{\mathcal{G}_{j(t)}}\right\rangle + \frac{L_{\max}}{2}\left\|(\Delta_t^s)_{\mathcal{G}_{j(t)}}\right\|^2 + g_{\mathcal{G}_{j(t)}}\left((x_{t+1}^s)_{\mathcal{G}_{j(t)}}\right) \\
& - g_{\mathcal{G}_{j(t)}}\left((x_t^s)_{\mathcal{G}_{j(t)}}\right) + \left\langle \nabla_{\mathcal{G}_{j(t)}} f(x_t^s) - (\widehat{v}_t^s)_{\mathcal{G}_{j(t)}}, (\Delta_t^s)_{\mathcal{G}_{j(t)}}\right\rangle\Bigg) \\
\leq\ & F(x_t^s) + \mathbb{E}_{j(t)}\Bigg(-\frac{L_{\max}}{\gamma}\left\|(\Delta_t^s)_{\mathcal{G}_{j(t)}}\right\|^2 + \frac{L_{\max}}{2}\left\|(\Delta_t^s)_{\mathcal{G}_{j(t)}}\right\|^2 + \\
& + \left\langle \nabla_{\mathcal{G}_{j(t)}} f(x_t^s) - (\widehat{v}_t^s)_{\mathcal{G}_{j(t)}}, (\Delta_t^s)_{\mathcal{G}_{j(t)}}\right\rangle\Bigg) \\
=\ & F(x_t^s) + \mathbb{E}_{j(t)}\left(\left(\frac{L_{\max}}{2} - \frac{L_{\max}}{\gamma}\right)\left\|(\Delta_t^s)_{\mathcal{G}_{j(t)}}\right\|^2\right) + \mathbb{E}_{j(t)}\left\langle \nabla_{\mathcal{G}_{j(t)}} f(x_t^s) - (\widehat{v}_t^s)_{\mathcal{G}_{j(t)}}, (\Delta_t^s)_{\mathcal{G}_{j(t)}}\right\rangle
\end{aligned}
$$

$$= \quad F(x_t^s) + \frac{L_{\max}}{k}\left(\frac{1}{2} - \frac{1}{\gamma}\right)\|\overline{x}_{t+1}^s - x_t^s\|^2 + \mathbb{E}_{j(t)}\left\langle \nabla_{\mathcal{G}_{j(t)}} f(x_t^s) - (\widehat{v}_t^s)_{\mathcal{G}_{j(t)}}, (\Delta_t^s)_{\mathcal{G}_{j(t)}} \right\rangle$$

where the first inequality uses (6), and the second inequality uses (29) in Lemma 2. Consider the expectation of the last term on the right-hand side of (45), we have

$$\begin{aligned}
& \mathbb{E}\left\langle \nabla_{\mathcal{G}_{j(t)}} f(x_t^s) - (\widehat{v}_t^s)_{\mathcal{G}_{j(t)}}, (\Delta_t^s)_{\mathcal{G}_{j(t)}} \right\rangle \qquad (46)\\
= \quad & \mathbb{E}\left\langle \nabla_{\mathcal{G}_{j(t)}} f(x_t^s) - (\nabla f_{i_t}(\widehat{x}_t^s) - \nabla f_{i_t}(\widetilde{x}^s) + \nabla f(\widetilde{x}^s))_{\mathcal{G}_{j(t)}}, (\Delta_t^s)_{\mathcal{G}_{j(t)}} \right\rangle \\
= \quad & \mathbb{E}\left\langle \nabla_{\mathcal{G}_{j(t)}} f(x_t^s) - (\nabla f_{i_t}(\widehat{x}_t^s) - \nabla f_{i_t}(x_t^s) + \nabla f_{i_t}(x_t^s) - \nabla f_{i_t}(\widetilde{x}^s) + \nabla f(\widetilde{x}^s))_{\mathcal{G}_{j(t)}}, (\Delta_t^s)_{\mathcal{G}_{j(t)}} \right\rangle \\
= \quad & \mathbb{E}\left\langle \nabla_{\mathcal{G}_{j(t)}} f(x_t^s) - \nabla_{\mathcal{G}_{j(t)}} f(\widetilde{x}^s), (\Delta_t^s)_{\mathcal{G}_{j(t)}} \right\rangle + \mathbb{E}\left\langle \nabla_{\mathcal{G}_{j(t)}} f_{i_t}(x_t^s) - \nabla_{\mathcal{G}_{j(t)}} f_{i_t}(\widehat{x}_t^s), (\Delta_t^s)_{\mathcal{G}_{j(t)}} \right\rangle \\
& + \mathbb{E}\left\langle \nabla_{\mathcal{G}_{j(t)}} f_{i_t}(\widetilde{x}^s) - \nabla_{\mathcal{G}_{j(t)}} f_{i_t}(x_t^s), (\Delta_t^s)_{\mathcal{G}_{j(t)}} \right\rangle \\
\leq \quad & \mathbb{E}\left(\left\|\nabla_{\mathcal{G}_{j(t)}} f(x_t^s) - \nabla_{\mathcal{G}_{j(t)}} f(\widetilde{x}^s)\right\| \|\Delta_t^s\|\right) + \mathbb{E}\left(\left\|\nabla_{\mathcal{G}_{j(t)}} f_{i_t}(x_t^s) - \nabla_{\mathcal{G}_{j(t)}} f_{i_t}(\widehat{x}_t^s)\right\| \|\Delta_t^s\|\right) \\
& + \mathbb{E}\left(\left\|\nabla_{\mathcal{G}_{j(t)}} f_{i_t}(\widetilde{x}^s) - \nabla_{\mathcal{G}_{j(t)}} f_{i_t}(x_t^s)\right\| \|\Delta_t^s\|\right) \\
= \quad & \frac{1}{k}\mathbb{E}\left(\|\nabla f(x_t^s) - \nabla f(\widetilde{x}^s)\| \left\|\overline{x}_{t+1}^s - x_t^s\right\|\right) + \frac{1}{k}\mathbb{E}\left(\|\nabla f_{i_t}(x_t^s) - \nabla f_{i_t}(\widehat{x}_t^s)\| \left\|\overline{x}_{t+1}^s - x_t^s\right\|\right) \\
& + \frac{1}{k}\mathbb{E}\left(\|\nabla f_{i_t}(\widetilde{x}^s) - \nabla f_{i_t}(x_t^s)\| \left\|\overline{x}_{t+1}^s - x_t^s\right\|\right) \\
\leq \quad & \frac{1}{k}\mathbb{E}\left(2L_{nor}\|x_t^s - \widetilde{x}^s\|\|\overline{x}_{t+1}^s - x_t^s\| + L_{res}\sum_{t' \in K(t)} \|\Delta_{t'}^s\| \left\|\overline{x}_{t+1}^s - x_t^s\right\|\right) \\
\leq \quad & \frac{1}{k}\mathbb{E}\left(2L_{nor}\sum_{t'=0}^{t-1} \|\Delta_{t'}^s\|\|\overline{x}_{t+1}^s - x_t^s\| + L_{res}\sum_{t'=t-\tau}^{t-1} \|\Delta_{t'}^s\|\|\overline{x}_{t+1}^s - x_t^s\|\right) \\
\leq \quad & 2L_{nor}\sum_{t'=0}^{t-1} \frac{\rho^{\frac{t-t'}{2}}}{k^{3/2}}\mathbb{E}\|\overline{x}_{t+1}^s - x_t^s\|^2 + L_{res}\sum_{t'=t-\tau}^{t-1} \frac{\rho^{\frac{t-t'}{2}}}{k^{3/2}}\mathbb{E}\|\overline{x}_{t+1}^s - x_t^s\|^2 \\
= \quad & k^{-3/2}\left(2L_{nor}\frac{\rho^{\frac{1}{2}} - \rho^{\frac{m}{2}}}{1-\rho^{\frac{1}{2}}} + L_{res}\frac{\rho^{\frac{1}{2}} - \rho^{\frac{\tau+1}{2}}}{1-\rho^{\frac{1}{2}}}\right) \cdot \mathbb{E}\|\overline{x}_{t+1}^s - x_t^s\|^2 \\
= \quad & k^{-3/2}\left(2L_{nor}\theta_2 + L_{res}\theta_1\right)\mathbb{E}\|\overline{x}_{t+1}^s - x_t^s\|^2
\end{aligned}$$

where the first inequality uses the Cauchy-Schwarz inequality (Callebaut, 1965), the third inequality uses (3) and (4), the sixth inequality uses (41).

By taking expectations on both sides of (45) and substituting (46), we have

$$\begin{aligned}
& \mathbb{E}F(x_{t+1}^s) \qquad (47)\\
\leq \quad & \mathbb{E}F(x_t^s) + \frac{L_{\max}}{k}\left(\frac{1}{2} - \frac{1}{\gamma}\right)\mathbb{E}\|\overline{x}_{t+1}^s - x_t^s\|^2 + \mathbb{E}\left\langle \nabla_{\mathcal{G}_{j(t)}} f(x_t^s) - (\widehat{v}_t^s)_{\mathcal{G}_{j(t)}}, (\Delta_t^s)_{\mathcal{G}_{j(t)}} \right\rangle \\
\leq \quad & \mathbb{E}F(x_t^s) - \frac{1}{k}\cdot\left(L_{\max}\left(\frac{1}{\gamma} - \frac{1}{2}\right) - \frac{2L_{nor}\theta_2 + L_{res}\theta_1}{k^{1/2}}\right)\mathbb{E}\|\overline{x}_{t+1}^s - x_t^s\|^2
\end{aligned}$$

where $L_{\max}\left(\frac{1}{\gamma} - \frac{1}{2}\right) - \frac{2L_{nor}\theta_2 + L_{res}\theta_1}{k^{1/2}} \geq 0$ because that $\gamma^{-1} \geq \frac{1}{2} + \frac{2\Lambda_{nor}\theta_2 + \Lambda_{res}\theta_1}{k^{1/2}}$. This completes the proof. ■

**Theorem 5** *Let $\rho$ be a constant that satisfies $\rho > 1$, and define the quantity $\theta' = \frac{\rho^{\tau+1}-\rho}{\rho-1}$. Suppose the nonnegative steplength parameter $\gamma > 0$ satisfies $1-\Lambda_{nor}\gamma-\frac{\gamma\tau\theta'}{n}-\frac{2(\Lambda_{res}\theta_1+\Lambda_{nor}\theta_2)\gamma}{n^{1/2}} \geq 0$. If the optimal strong convexity holds for $f$ with $l > 0$, we have*

$$\mathbb{E}F(x^s) - F^* \leq \frac{L_{\max}}{2\gamma}\left(\frac{1}{1+\frac{m\gamma l}{k(l\gamma+L_{\max})}}\right)^s \cdot \left(\|x^0 - \mathcal{P}_S(x^0)\|^2 + \frac{2\gamma}{L_{\max}}\left(\mathbb{E}F(x^0) - F^*\right)\right) \quad (48)$$

*If $f$ is a general smooth convex function, we have*

$$\mathbb{E}F(x^s) - F^* \leq \frac{kL_{\max}\|x^0 - \mathcal{P}_S(x^0)\|^2 + 2\gamma k\left(F(x^0) - F^*\right)}{2\gamma k + 2m\gamma s} \quad (49)$$

**Proof** We have that

$$\begin{aligned}
& \|x^s_{t+1} - \mathcal{P}_S(x^s_{t+1})\|^2 \leq \|x^s_{t+1} - \mathcal{P}_S(x^s_t)\|^2 = \|x^s_t + \Delta^s_t - \mathcal{P}_S(x^s_t)\|^2 \quad (50)\\
= & \|x^s_t - \mathcal{P}_S(x^s_t)\|^2 - \|\Delta^s_t\|^2 - 2\langle(\mathcal{P}_S(x^s_t) - x^s_t - \Delta^s_t)_{\mathcal{G}_{j(t)}}, (\Delta^s_t)_{\mathcal{G}_{j(t)}}\rangle \\
= & \|x^s_t - \mathcal{P}_S(x^s_t)\|^2 - \|\Delta^s_t\|^2 - 2\langle\left(\mathcal{P}_S(x^s_t) - x^s_{t+1}\right)_{\mathcal{G}_{j(t)}}, (\Delta^s_t)_{\mathcal{G}_{j(t)}}\rangle \\
\leq & \|x^s_t - \mathcal{P}_S(x^s_t)\|^2 - \|\Delta^s_t\|^2 + \frac{2\gamma}{L_{\max}}\left(\langle\left(\mathcal{P}_S(x^s_t) - x^s_{t+1}\right)_{\mathcal{G}_{j(t)}}, (\widehat{v}^s_t)_{\mathcal{G}_{j(t)}}\rangle\right) + \\
& \frac{2\gamma}{L_{\max}}\left(g_{\mathcal{G}_{j(t)}}(\mathcal{P}_S(x^s_t)_{\mathcal{G}_{j(t)}}) - g_{j(t)}(x^s_{t+1})_{\mathcal{G}_{j(t)}})\right) \\
= & \|x^s_t - \mathcal{P}_S(x^s_t)\|^2 - \|\Delta^s_t\|^2 + \frac{2\gamma}{L_{\max}}\underbrace{\left(\langle(\mathcal{P}_S(x^s_t) - x^s_t)_{\mathcal{G}_{j(t)}}, (\widehat{v}^s_t)_{\mathcal{G}_{j(t)}}\rangle\right)}_{T_1} + \\
& \frac{2\gamma}{L_{\max}}\underbrace{\left(\langle(\Delta^s_t)_{\mathcal{G}_{j(t)}}, (\widehat{v}^s_t)_{\mathcal{G}_{j(t)}}\rangle\right)}_{T_2} + \frac{2\gamma}{L_{\max}}\left(g_{j(t)}(\mathcal{P}_S(x^s_t)_{\mathcal{G}_{j(t)}}) - g_{\mathcal{G}_{j(t)}}(x^s_{t+1})_{\mathcal{G}_{j(t)}})\right)
\end{aligned}$$

where the first inequality comes from the definition of function $\mathcal{P}_{S(x)} = \arg\min_{y\in S}\|y-x\|^2$, and the second inequality uses (29) in Lemma 2. For the expectation of $T_1$, we have

$$\begin{aligned}
& \mathbb{E}(T_1) = \mathbb{E}\left(\langle(\mathcal{P}_S(x^s_t) - x^s_t)_{\mathcal{G}_{j(t)}}, (\widehat{v}^s_t)_{\mathcal{G}_{j(t)}}\rangle\right) \quad (51)\\
= & \frac{1}{k}\mathbb{E}\langle\mathcal{P}_S(x^s_t) - x^s_t, \widehat{v}^s_t\rangle \\
= & \frac{1}{k}\mathbb{E}\langle\mathcal{P}_S(x^s_t) - x^s_t, \nabla f_{i_t}(\widehat{x}^s_t) - \nabla f_{i_t}(\widetilde{x}^s) + \nabla f(\widetilde{x}^s)\rangle \\
= & \frac{1}{k}\mathbb{E}\langle\mathcal{P}_S(x^s_t) - x^s_t, \nabla f_{i_t}(\widehat{x}^s_t)\rangle + \frac{1}{k}\langle\mathbb{E}(\mathcal{P}_S(x^s_t) - x^s_t), \mathbb{E}(-\nabla f_{i_t}(\widetilde{x}^s) + \nabla f(\widetilde{x}^s))\rangle \\
= & \frac{1}{k}\mathbb{E}\langle\mathcal{P}_S(x^s_t) - \widehat{x}^s_t, \nabla f_{i_t}(\widehat{x}^s_t)\rangle + \frac{1}{k}\mathbb{E}\langle\widehat{x}^s_t - x^s_t, \nabla f_{i_t}(\widehat{x}^s_t)\rangle \\
= & \frac{1}{k}\mathbb{E}\langle\mathcal{P}_S(x^s_t) - \widehat{x}^s_t, \nabla f_{i_t}(\widehat{x}^s_t)\rangle + \frac{1}{k}\mathbb{E}\sum_{t'=0}^{|K(t)|-1}\langle\widehat{x}^s_{t,t'} - \widehat{x}^s_{t,t'+1}, \nabla f_{i_t}(\widehat{x}^s_t)\rangle
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{k}\mathbb{E}\langle \mathcal{P}_S(x_t^s) - \widehat{x}_t^s, \nabla f_{i_t}(\widehat{x}_t^s)\rangle + \frac{1}{k}\mathbb{E}\sum_{t'=0}^{|K(t)|-1}\langle \widehat{x}_{t,t'}^s - \widehat{x}_{t,t'+1}^s, \nabla f_{i_t}(\widehat{x}_{t,t'}^s)\rangle \\
&\quad + \frac{1}{k}\mathbb{E}\sum_{t'=0}^{|K(t)|-1}\langle \widehat{x}_{t,t'}^s - \widehat{x}_{t,t'+1}^s, \nabla f_{i_t}(\widehat{x}_t^s) - \nabla f_{i_t}(\widehat{x}_{t,t'}^s)\rangle \\
&\leq \frac{1}{k}\mathbb{E}\left(f_{i_t}(\mathcal{P}_S(x_t^s)) - f_{i_t}(\widehat{x}_t^s)\right) + \frac{1}{k}\mathbb{E}\sum_{t'=0}^{|K(t)|-1}\langle \widehat{x}_{t,t'}^s - \widehat{x}_{t,t'+1}^s, \nabla f_{i_t}(\widehat{x}_t^s) - \nabla f_{i_t}(\widehat{x}_{t,t'}^s)\rangle \\
&\quad + \frac{1}{k}\mathbb{E}\sum_{t'=0}^{|K(t)|-1}\left(f_{i_t}(\widehat{x}_{t,t'}^s) - f_{i_t}(\widehat{x}_{t,t'+1}^s) + \frac{L_{\max}}{2}\|\widehat{x}_{t,t'}^s - \widehat{x}_{t,t'+1}^s\|^2\right) \\
&= \frac{1}{k}\mathbb{E}\left(f_{i_t}(\mathcal{P}_S(x_t^s)) - f_{i_t}(x_t^s)\right) + \frac{1}{k}\mathbb{E}\sum_{t'=0}^{|K(t)|-1}\langle \widehat{x}_{t,t'}^s - \widehat{x}_{t,t'+1}^s, \nabla f_{i_t}(\widehat{x}_t^s) - \nabla f_{i_t}(\widehat{x}_{t,t'}^s)\rangle \\
&\quad + \frac{L_{\max}}{2k}\mathbb{E}\sum_{t'=0}^{|K(t)|-1}\left(\|\widehat{x}_{t,t'}^s - \widehat{x}_{t,t'+1}^s\|^2\right) \\
&= \frac{1}{k}\mathbb{E}\left(f_{i_t}(\mathcal{P}_S(x_t^s)) - f_{i_t}(x_t^s)\right) + \frac{L_{\max}}{2k}\mathbb{E}\sum_{t'=0}^{|K(t)|-1}\left(\|\widehat{x}_{t,t'}^s - \widehat{x}_{t,t'+1}^s\|^2\right) \\
&\quad + \frac{1}{k}\mathbb{E}\sum_{t'=0}^{|K(t)|-1}\langle \widehat{x}_{t,t'}^s - \widehat{x}_{t,t'+1}^s, \sum_{t''=0}^{t'-1}\nabla f_{i_t}(\widehat{x}_{t''}^s) - \nabla f_{i_t}(\widehat{x}_{t,t''+1}^s)\rangle \\
&\leq \frac{1}{k}\mathbb{E}\left(f_{i_t}(\mathcal{P}_S(x_t^s)) - f_{i_t}(x_t^s)\right) + \frac{L_{\max}}{2k}\mathbb{E}\sum_{t'=0}^{|K(t)|-1}\left(\|\widehat{x}_{t,t'}^s - \widehat{x}_{t,t'+1}^s\|^2\right) \\
&\quad + \frac{L_{\max}}{k}\mathbb{E}\sum_{t'=0}^{|K(t)|-1}\left(\left\|\widehat{x}_{t,t'}^s - \widehat{x}_{t,t'+1}^s\right\|\sum_{t''=0}^{t'-1}\left\|\widehat{x}_{t''}^s - \widehat{x}_{t,t''+1}^s\right\|\right) \\
&= \frac{1}{k}\mathbb{E}\left(f_{i_t}(\mathcal{P}_S(x_t^s)) - f_{i_t}(x_t^s)\right) + \frac{L_{\max}}{2k}\mathbb{E}\sum_{t'=0}^{|K(t)|-1}\left(\|\widehat{x}_{t,t'}^s - \widehat{x}_{t,t'+1}^s\|^2\right) \\
&\quad + \frac{L_{\max}}{2k}\mathbb{E}\left(\left(\sum_{t'=0}^{|K(t)|-1}\left\|\widehat{x}_{t,t'}^s - \widehat{x}_{t,t'+1}^s\right\|\right)^2 - \sum_{t'=0}^{|K(t)|-1}\left(\|\widehat{x}_{t,t'}^s - \widehat{x}_{t,t'+1}^s\|^2\right)\right) \\
&= \frac{1}{k}\mathbb{E}\left(f_{i_t}(\mathcal{P}_S(x_t^s)) - f_{i_t}(x_t^s)\right) + \frac{L_{\max}}{2k}\mathbb{E}\left(\sum_{t'=0}^{|K(t)|-1}\left\|\widehat{x}_{t,t'}^s - \widehat{x}_{t,t'+1}^s\right\|\right)^2 \\
&\leq \frac{1}{k}\mathbb{E}\left(f_{i_t}(\mathcal{P}_S(x_t^s)) - f_{i_t}(x_t^s)\right) + \frac{L_{\max}\tau}{2k}\mathbb{E}\sum_{t'=0}^{|K(t)|-1}\left\|\widehat{x}_{t,t'}^s - \widehat{x}_{t,t'+1}^s\right\|^2 \\
&= \frac{1}{k}\mathbb{E}\left(f_{i_t}(\mathcal{P}_S(x_t^s)) - f_{i_t}(x_t^s)\right) + \frac{L_{\max}\tau}{2k}\sum_{t'=0}^{|K(t)|-1}\mathbb{E}\left\|B_{t'}^s\Delta_{t'}^s\right\|^2
\end{aligned}
$$

$$
\begin{aligned}
&\leq \frac{1}{k}\mathbb{E}\left(f_{i_t}(\mathcal{P}_S(x_t^s)) - f_{i_t}(x_t^s)\right) + \frac{L_{\max}\tau}{2k^2}\sum_{t'=0}^{|K(t)|-1}\mathbb{E}\left\|\overline{x}_{t'+1}^s - x_{t'}^s\right\|^2 \\
&\leq \frac{1}{k}\mathbb{E}\left(f_{i_t}(\mathcal{P}_S(x_t^s)) - f_{i_t}(x_t^s)\right) + \frac{L_{\max}\tau}{2k^2}\sum_{t'=1}^{\tau}\rho^{t'}\mathbb{E}\left\|\overline{x}_{t+1}^s - x_t^s\right\|^2 \\
&\leq \frac{1}{k}\mathbb{E}\left(f(\mathcal{P}_S(x_t^s)) - f(x_t^s)\right) + \frac{L_{\max}\tau\theta'}{2k^2}\mathbb{E}(\|x_t^s - \overline{x}_{t+1}^s\|^2)
\end{aligned}
$$

where the fifth equality comes from that $x_t^s$ is independent to $i_t$, the sixth equality uses Lemma 1, the first inequality uses the convexity of $f_i$ and (6), the second inequality uses (5). For the expectation of $T_2$, we have

$$
\begin{aligned}
&\mathbb{E}(T_2) = \mathbb{E}\langle(\Delta_t^s)_{\mathcal{G}_{j(t)}}, (\widehat{v}_t^s)_{\mathcal{G}_{j(t)}}\rangle \qquad (52)\\
=\;& \mathbb{E}\langle(\Delta_t)_{\mathcal{G}_{j(t)}}, (\nabla f_{i_t}(\widehat{x}_t^s) - \nabla f_{i_t}(\widetilde{x}^s) + \nabla f(\widetilde{x}^s))_{\mathcal{G}_{j(t)}}\rangle \\
=\;& \mathbb{E}\langle(\Delta_t)_{\mathcal{G}_{j(t)}}, (\nabla f_{i_t}(\widehat{x}_t^s) - \nabla f_{i_t}(x_t^s) + \nabla f_{i_t}(x_t^s) - \nabla f_{i_t}(\widetilde{x}^s) + \nabla f(\widetilde{x}^s))_{\mathcal{G}_{j(t)}}\rangle \\
=\;& \mathbb{E}\langle(\Delta_t)_{\mathcal{G}_{j(t)}}, (\nabla f_{i_t}(\widehat{x}_t^s) - \nabla f_{i_t}(x_t^s))_{\mathcal{G}_{j(t)}}\rangle + \\
& \mathbb{E}\langle(\Delta_t)_{\mathcal{G}_{j(t)}}, (\nabla f_{i_t}(x_t^s) - \nabla f_{i_t}(\widetilde{x}^s))_{\mathcal{G}_{j(t)}}\rangle + \mathbb{E}\langle(\Delta_t)_{\mathcal{G}_{j(t)}}, \nabla_{\mathcal{G}_{j(t)}} f(\widetilde{x}^s)\rangle \\
\leq\;& \frac{1}{k}\mathbb{E}\left(\|\overline{x}_{t+1}^s - x_t^s\|\,\|\nabla f_{i_t}(\widehat{x}_t^s) - \nabla f_{i_t}(x_t^s)\|\right) + \\
& \frac{1}{k}\mathbb{E}\left(\|\overline{x}_{t+1}^s - x_t^s\|\,\|\nabla f_{i_t}(x_t^s) - \nabla f_{i_t}(\widetilde{x}^s)\|\right) + \mathbb{E}\langle(\Delta_t)_{\mathcal{G}_{j(t)}}, \nabla_{\mathcal{G}_{j(t)}} f(\widetilde{x}^s)\rangle \\
\leq\;& \frac{L_{res}}{k}\left(\sum_{t'\in K(t)}\|\overline{x}_{t+1}^s - x_t^s\|\|\Delta_{t'}^s\|\right) \\
& \frac{L_{nor}}{k}\mathbb{E}\left(\|\overline{x}_{t+1}^s - x_t^s\|\|x_t^s - x^s\|\right) + \mathbb{E}\langle(\Delta_t)_{\mathcal{G}_{j(t)}}, \nabla_{\mathcal{G}_{j(t)}} f(\widetilde{x}^s)\rangle \\
\leq\;& \frac{L_{res}}{k}\left(\sum_{t'\in K(t)}\|\overline{x}_{t+1}^s - x_t^s\|\|\Delta_{t'}^s\|\right) \\
& \frac{L_{nor}}{k}\mathbb{E}\left(\sum_{t'=0}^{t-1}\|\overline{x}_{t+1}^s - x_t^s\|\|\Delta_{t'}^s\|\right) + \mathbb{E}\langle(\Delta_t)_{\mathcal{G}_{j(t)}}, \nabla_{\mathcal{G}_{j(t)}} f(\widetilde{x}^s)\rangle \\
\leq\;& \frac{L_{res}}{k^{3/2}}\sum_{t'=t-\tau}^{t-1}\rho^{(t-t')/2}\mathbb{E}(\|x_t^s - \overline{x}_{t+1}^s\|^2) \\
& + \frac{L_{nor}}{k^{3/2}}\sum_{t'=0}^{t-1}\rho^{(t-t')/2}\mathbb{E}(\|x_t^s - \overline{x}_{t+1}^s\|^2) + \mathbb{E}\langle(\Delta_t)_{\mathcal{G}_{j(t)}}, \nabla_{\mathcal{G}_{j(t)}} f(\widetilde{x}^s)\rangle \\
=\;& \frac{1}{k^{3/2}}\left(L_{res}\theta_1 + L_{nor}\theta_2\right)\mathbb{E}(\|x_t^s - \overline{x}_{t+1}^s\|^2) + \mathbb{E}\langle(\Delta_t)_{\mathcal{G}_{j(t)}}, \nabla_{\mathcal{G}_{j(t)}} f(\widetilde{x}^s)\rangle \\
\leq\;& \frac{L_{res}\theta_1 + L_{nor}\theta_2}{k^{3/2}}\mathbb{E}(\|x_t^s - \overline{x}_{t+1}^s\|^2) + \mathbb{E}\langle(\Delta_t)_{\mathcal{G}_{j(t)}}, \nabla_{\mathcal{G}_{j(t)}} f(\widetilde{x}^s)\rangle
\end{aligned}
$$

where the second inequality uses Lemma 1, the fourth inequality uses (41). By substituting the upper bounds from (51) and (52) into (50), we have

$$
\begin{aligned}
& \mathbb{E}\|x_{t+1}^s - \mathcal{P}_S(x_{t+1}^s)\|^2 \qquad (53)\\
\leq\ & \mathbb{E}\|x_t^s - \mathcal{P}_S(x_t^s)\|^2 - \frac{1}{k}\mathbb{E}(\|x_t^s - \overline{x}_{t+1}^s\|^2) + \\
& \frac{2\gamma}{L_{\max}k}\mathbb{E}\left(f(\mathcal{P}_S(x_t^s)) - f(x_t^s)\right) + \frac{\gamma\tau\theta'}{k^2}\mathbb{E}(\|x_t^s - \overline{x}_{t+1}^s\|^2) \\
& + \frac{2\gamma}{L_{\max}}\Bigg(\frac{L_{res}\theta_1 + L_{nor}\theta_2}{k^{3/2}}\theta\mathbb{E}(\|x_t^s - \overline{x}_{t+1}^s\|^2) + \mathbb{E}\langle(\Delta_t)_{\mathcal{G}_{j(t)}}, \nabla_{\mathcal{G}_{j(t)}} f(\widetilde{x}^s)\rangle \\
& \frac{1}{k}\mathbb{E}g(\mathcal{P}_S(x_t^s)) - \mathbb{E}g(x_{t+1}^s) + \frac{k-1}{k}\mathbb{E}g(x_t^s)\Bigg) \\
=\ & \mathbb{E}\|x_t^s - \mathcal{P}_S(x_t^s)\|^2 + \frac{2\gamma}{L_{\max}k}\mathbb{E}\left(f(\mathcal{P}_S(x_t^s)) - f(x_t^s)\right) - \\
& \frac{1}{k}\left(1 - \frac{\gamma\tau\theta'}{k} - \frac{2(L_{res}\theta_1 + L_{nor}\theta_2)\gamma}{k^{1/2}L_{\max}}\right)\mathbb{E}(\|x_t^s - \overline{x}_{t+1}^s\|^2) + \frac{2\gamma}{L_{\max}}\Bigg(\mathbb{E}\langle(\Delta_t)_{\mathcal{G}_{j(t)}}, \nabla_{\mathcal{G}_{j(t)}} f(\widetilde{x}^s)\rangle \\
& \frac{1}{k}\mathbb{E}g(\mathcal{P}_S(x_t^s)) - \mathbb{E}g(x_{t+1}^s) + \frac{k-1}{k}\mathbb{E}g(x_t^s)\Bigg)
\end{aligned}
$$

We consider a fixed stage $s+1$ such that $x_0^{s+1} = x_m^s$. By summing the the inequality (53) over $t = 0, \cdots, m-1$, we obtain

$$
\begin{aligned}
& \mathbb{E}\|x^{s+1} - \mathcal{P}_S(x^{s+1})\|^2 \qquad (54)\\
\leq\ & \mathbb{E}\|x^s - \mathcal{P}_S(x^s)\|^2 + \sum_{t'=0}^{m-1}\frac{2\gamma}{L_{\max}k}\mathbb{E}\left(f(\mathcal{P}_S(x_{t'}^{s+1})) - f(x_{t'}^{s+1})\right) - \\
& \sum_{t'=0}^{m-1}\frac{1}{k}\left(1 - \frac{\gamma\tau\theta'}{k} - \frac{2(L_{res}\theta_1 + L_{nor}\theta_2)\gamma}{k^{1/2}L_{\max}}\right)\cdot\mathbb{E}(\|x_{t'}^{s+1} - \overline{x}_{t'+1}^{s+1}\|^2) \\
& + \frac{2\gamma}{L_{\max}}\sum_{t'=0}^{m-1}\mathbb{E}\left\langle(\Delta_{t'})_{\mathcal{G}_{j(t')}}, \nabla_{\mathcal{G}_{j(t')}} f(\widetilde{x}^s)\right\rangle \\
& + \frac{2\gamma}{L_{\max}}\sum_{t'=0}^{m-1}\left(\frac{1}{k}\mathbb{E}g(\mathcal{P}_S(x_{t'}^{s+1})) - \mathbb{E}g(x_{t'+1}^{s+1}) + \frac{n-1}{k}\mathbb{E}g(x_{t'}^{s+1})\right) \\
=\ & \mathbb{E}\|x^s - \mathcal{P}_S(x^s)\|^2 + \sum_{t'=0}^{m-1}\frac{2\gamma}{L_{\max}k}\mathbb{E}\left(f(\mathcal{P}_S(x_{t'}^{s+1})) - f(x_{t'}^{s+1})\right) - \\
& \sum_{t'=0}^{m-1}\frac{1}{k}\left(1 - \frac{\gamma\tau\theta'}{k} - \frac{2(L_{res}\theta_1 + L_{nor}\theta_2)\gamma}{k^{1/2}L_{\max}}\right)\cdot\mathbb{E}(\|x_{t'}^{s+1} - \overline{x}_{t'+1}^{s+1}\|^2) \\
& + \frac{2\gamma}{L_{\max}}\mathbb{E}\left\langle x^{s+1} - x^s, \nabla f(\widetilde{x}^s)\right\rangle \\
& + \frac{2\gamma}{L_{\max}}\sum_{t'=0}^{m-1}\left(\frac{1}{k}\mathbb{E}g(\mathcal{P}_S(x_{t'}^{s+1})) - \mathbb{E}g(x_{t'+1}^{s+1}) + \frac{k-1}{k}\mathbb{E}g(x_{t'}^{s+1})\right)
\end{aligned}
$$

$$
\begin{aligned}
\leq\ & \mathbb{E}\|x^s-\mathcal{P}_S(x^s)\|^2+\sum_{t'=0}^{m-1}\frac{2\gamma}{L_{\max}k}\mathbb{E}\left(f(\mathcal{P}_S(x_{t'}^{s+1}))-f(x_{t'}^{s+1})\right)- \\
& \sum_{t'=0}^{m-1}\frac{1}{k}\left(1-\frac{\gamma\tau\theta'}{k}-\frac{2(L_{res}\theta_1+L_{nor}\theta_2)\gamma}{k^{1/2}L_{\max}}\right)\cdot\mathbb{E}(\|x_{t'}^{s+1}-\overline{x}_{t'+1}^{s+1}\|^2) \\
& +\frac{2\gamma}{L_{\max}}\mathbb{E}\left(f(x^s)-f(x^{s+1})+\frac{L_{nor}}{2}\|x^{s+1}-x^s\|^2\right) \\
& +\frac{2\gamma}{L_{\max}}\sum_{t'=0}^{m-1}\left(\frac{1}{k}\mathbb{E}g(\mathcal{P}_S(x_{t'}^{s+1}))-\mathbb{E}g(x_{t'+1}^{s+1})+\frac{k-1}{k}\mathbb{E}g(x_{t'}^{s+1})\right) \\
=\ & \mathbb{E}\|x^s-\mathcal{P}_S(x^s)\|^2+\sum_{t'=0}^{m-1}\frac{2\gamma}{L_{\max}k}\mathbb{E}\left(f(\mathcal{P}_S(x_{t'}^{s+1}))-f(x_{t'}^{s+1})\right)- \\
& \sum_{t'=0}^{m-1}\frac{1}{k}\left(1-\frac{\gamma\tau\theta'}{k}-\frac{2(L_{res}\theta_1+L_{nor}\theta_2)\gamma}{k^{1/2}L_{\max}}\right)\cdot\mathbb{E}(\|x_{t'}^{s+1}-\overline{x}_{t'+1}^{s+1}\|^2) \\
& +\frac{2\gamma}{L_{\max}}\sum_{t'=0}^{m-1}\mathbb{E}\left(f(x_{t'}^{s+1})-f(x_{t+1}^{s+1})\right)+\frac{L_{nor}\gamma}{L_{\max}}\mathbb{E}\left\|\sum_{t'=0}^{m-1}\left(x_{t'}^{s+1}-x_{t'+1}^{s+1}\right)\right\|^2 \\
& +\frac{2\gamma}{L_{\max}}\sum_{t'=0}^{m-1}\left(\frac{1}{k}\mathbb{E}g(\mathcal{P}_S(x_{t'}^{s+1}))-\mathbb{E}g(x_{t'+1}^{s+1})+\frac{k-1}{k}\mathbb{E}g(x_{t'}^{s+1})\right) \\
\leq\ & \mathbb{E}\|x^s-\mathcal{P}_S(x^s)\|^2+\frac{2\gamma}{L_{\max}k}\sum_{t'=0}^{m-1}\left(F^*-\mathbb{E}F(x_{t'}^{s+1})\right)+\frac{2\gamma}{L_{\max}}\sum_{t'=0}^{m-1}\left(\mathbb{E}F(x_{t'}^{s+1})-\mathbb{E}F(x_{t'+1}^{s+1})\right) \\
& -\sum_{t'=0}^{m-1}\frac{1}{k}\left(1-\Lambda_{nor}\gamma-\frac{\gamma\tau\theta'}{n}-\frac{2(\Lambda_{res}\theta_1+\Lambda_{nor}\theta_2)\gamma}{k^{1/2}}\right)\cdot\mathbb{E}(\|x_{t'}^{s+1}-\overline{x}_{t'+1}^{s+1}\|^2) \\
\leq\ & \mathbb{E}\|x^s-\mathcal{P}_S(x^s)\|^2+\frac{2\gamma}{L_{\max}k}\sum_{t'=0}^{m-1}\left(F^*-\mathbb{E}F(x_{t'}^{s+1})\right)+\frac{2\gamma}{L_{\max}}\left(\mathbb{E}F(x^s)-\mathbb{E}F(x^{s+1})\right)
\end{aligned}
$$

where the second inequality uses (3), the final inequality comes from $1-\Lambda_{nor}\gamma-\frac{\gamma\tau\theta'}{n}-\frac{2(\Lambda_{res}\theta_1+\Lambda_{nor}\theta_2)\gamma}{n^{1/2}}\geq 0$. Define $S(x^s)=\mathbb{E}\|x_t-\mathcal{P}_S(x^s)\|^2+\frac{2\gamma}{L_{\max}}\mathbb{E}\left(F(x^s)-F^*\right)$. According to (54), we have

$$
S(x^{s+1})\leq S(x^s)-\frac{2\gamma}{L_{\max}k}\sum_{t'=0}^{m-1}\mathbb{E}\left(F(x_{t'}^{s+1})-F^*\right)\leq S(x^s)-\frac{2m\gamma}{L_{\max}k}\mathbb{E}\left(F(x^{s+1})-F^*\right) \quad (55)
$$

where the second inequality comes from the monotonicity of $\mathbb{E}F(x_t^s)$. According to (55), we have

$$
S(x^s)\leq S(x^0)-\frac{2m\gamma s}{L_{\max}k}\mathbb{E}\left(F(x^s)-F^*\right) \quad (56)
$$

Thus, the sublinear convergence rate (49) for general smooth convex function $f$ can be obtained from (56).

If the optimal strong convexity for the smooth convex function $f$ holds with $l > 0$, we have (57) as proved in (A.28) of Liu and Wright (2015).

$$\mathbb{E}\left(F(x^s) - F^*\right) \geq \frac{L_{\max} l}{2(l\gamma + L_{\max})} S(x^s) \tag{57}$$

Thus, substituting (58) into (55), we have

$$S(x^{s+1}) \leq S(x^s) - \frac{2m\gamma l}{2k(l\gamma + L_{\max})} S(x^{s+1}) \tag{58}$$

Based on (58), we have (59) by induction.

$$S(x^s) \leq \left( \frac{1}{1 + \frac{2m\gamma l}{2k(l\gamma + L_{\max})}} \right)^s S(x^0) \tag{59}$$

Thus, the linear convergence rate (48) for the optimal strong convexity on $f$ can be obtained from (59). This completes the proof. ■

**Theorem 6** *Let $\rho$ be a constant that satisfies $\rho > 1$, and define the quantities $\theta_1 = \frac{\rho^{\frac{1}{2}} - \rho^{\frac{\tau+1}{2}}}{1-\rho^{\frac{1}{2}}}$ and $\theta_2 = \frac{\rho^{\frac{1}{2}} - \rho^{\frac{m}{2}}}{1-\rho^{\frac{1}{2}}}$. Suppose the nonnegative steplength parameter $\gamma > 0$ satisfies $\gamma \leq \min\left\{ \frac{k^{1/2}(1-\rho^{-1})-4}{4(\Lambda_{res}(1+\theta_1)+\Lambda_{nor}(1+\theta_2))}, \frac{k^{1/2}}{\frac{1}{2}k^{1/2}+2\Lambda_{nor}\theta_2+\Lambda_{res}\theta_1} \right\}$. Let $T$ denote the number of total iterations of AsyDSPG+. If $f$ is a smooth non-convex function, we have*

$$\frac{1}{T}\sum_{s=0}^{S-1}\sum_{t=0}^{m-1} \mathbb{E}\left\| \widetilde{\nabla} F(x_t^s) \right\|^2 \leq \left( \frac{\gamma}{k}\left( \frac{1}{\gamma} - \frac{1}{2} - \frac{2L_{nor}\theta_2 + L_{res}\theta_1}{k^{1/2}L_{\max}} \right) \right)^{-1} \frac{F(x^0) - F^*}{T} \tag{60}$$

**Proof** According to (47), we have that

$$\frac{\gamma}{k} \cdot \left( \frac{1}{\gamma} - \frac{1}{2} - \frac{2L_{nor}\theta_2 + L_{res}\theta_1}{k^{1/2}L_{\max}} \right) \mathbb{E}\|\widetilde{\nabla} F(x_t^s)\|^2 \leq \mathbb{E}F(x_t^s) - \mathbb{E}F(x_{t+1}^s) \tag{61}$$

Because $\gamma \leq \min\left\{ \frac{k^{1/2}(1-\rho^{-1})-4}{4(\Lambda_{res}(1+\theta_1)+\Lambda_{nor}(1+\theta_2))}, \frac{k^{1/2}}{\frac{1}{2}k^{1/2}+2\Lambda_{nor}\theta_2+\Lambda_{res}\theta_1} \right\}$, we have that

$$\mathbb{E}\|\widetilde{\nabla} F(x_t^s)\|^2 \leq \left( \frac{\gamma}{k} \cdot \left( \frac{1}{\gamma} - \frac{1}{2} - \frac{2L_{nor}\theta_2 + L_{res}\theta_1}{k^{1/2}L_{\max}} \right) \right)^{-1} \left( \mathbb{E}F(x_t^s) - \mathbb{E}F(x_{t+1}^s) \right) \tag{62}$$

Combining the inequalities (63) for all iterations in AsyDSPG+, we have

$$\sum_{s=0}^{S-1}\sum_{t=0}^{m-1} \mathbb{E}\left\| \widetilde{\nabla} F(x_t^s) \right\|^2 \leq \left( \frac{\gamma}{k} \cdot \left( \frac{1}{\gamma} - \frac{1}{2} - \frac{2L_{nor}\theta_2 + L_{res}\theta_1}{k^{1/2}L_{\max}} \right) \right)^{-1} \left( F(x^0) - \mathbb{E}F(x^S) \right) \tag{63}$$

Based on (63), we have that

$$\frac{1}{T}\sum_{s=0}^{S-1}\sum_{t=0}^{m-1} \mathbb{E}\left\| \widetilde{\nabla} F(x_t^s) \right\|^2 \leq \left( \frac{\gamma}{k}\left( \frac{1}{\gamma} - \frac{1}{2} - \frac{2L_{nor}\theta_2 + L_{res}\theta_1}{k^{1/2}L_{\max}} \right) \right)^{-1} \frac{F(x^0) - F^*}{T} \tag{64}$$

This completes the proof. ■

**Remark 7** *If we constraint $|\mathcal{G}_j| = 1$ for all $j$ in (1), our AsyDSPG+ degenerates to the asynchronous stochastic proximal coordinate descent algorithm with variance reduction. Based on Theorem 4, we have the linear convergence rate (65) if the optimal strong convexity holds for $f$ with $l > 0$.*

$$\mathbb{E}F(x^s) - F^* \leq \frac{L_{\max}}{2\gamma}\left(\frac{1}{1+\frac{m\gamma l}{n(l\gamma+L_{\max})}}\right)^s \cdot \left(\|x^0 - \mathcal{P}_S(x^0)\|^2 + \frac{2\gamma}{L_{\max}}\left(\mathbb{E}F(x^0) - F^*\right)\right) \quad (65)$$

*If $f$ is a general smooth convex function, we have*

$$\mathbb{E}F(x^s) - F^* \leq \frac{nL_{\max}\|x^0 - \mathcal{P}_S(x^0)\|^2 + 2\gamma n\left(F(x^0) - F^*\right)}{2\gamma n + 2m\gamma s} \quad (66)$$

*If $f$ is a general smooth non-convex function, we have*

$$\frac{1}{T}\sum_{s=0}^{S-1}\sum_{t=0}^{m-1}\mathbb{E}\left\|\widetilde{\nabla}F(x_t^s)\right\|^2 \leq \left(\frac{\gamma}{n}\left(\frac{1}{\gamma} - \frac{1}{2} - \frac{2L_{nor}\theta_2 + L_{res}\theta_1}{n^{1/2}L_{\max}}\right)\right)^{-1}\frac{F(x^0) - F^*}{T} \quad (67)$$

## References


DK Callebaut. Generalization of the cauchy-schwarz inequality. *Journal of mathematical analysis and applications*, 12(3):491–494, 1965.

Ji Liu and Stephen J Wright. Asynchronous stochastic coordinate descent: Parallelism and convergence properties. *SIAM Journal on Optimization*, 25(1):351–376, 2015.

Horia Mania, Xinghao Pan, Dimitris Papailiopoulos, Benjamin Recht, Kannan Ramchandran, and Michael I Jordan. Perturbed iterate analysis for asynchronous stochastic optimization. *arXiv preprint arXiv:1507.06970*, 2015.

Meisam Razaviyayn, Mingyi Hong, Zhi-Quan Luo, and Jong-Shi Pang. Parallel successive convex approximation for nonsmooth nonconvex optimization. In *Advances in Neural Information Processing Systems*, pages 1440–1448, 2014.

Andrzej P Ruszczyński. *Nonlinear Optimization*. Number 13 in Nonlinear optimization. Princeton University Press, 2006.

Xiyu Yu and Dacheng Tao. Variance-reduced proximal stochastic gradient descent for non-convex composite optimization. *arXiv preprint arXiv:1606.00602*, 2016.