Supplemental Material of Paper "Adaptive Proximal Average based Variance Reducing Stochastic Methods for Optimization with Composite Regularization"

1 Proof of Lemma 2.

In this section, we give the proof of Lemma 2. In the s-th outer loop, we denote $\hat{x}_s^*$ as the minimum of the approximated function $\hat{F}$, x_s^l as x in the l-th inner loop. Besides, we adopt $\gamma_s = \rho^s/3L$ and $c_s = \frac{3L}{2(1-\mu\gamma_s)n}$ as stated in the main body.

Lemma 2. *Suppose that Assumptions 1, 2 and 3 hold and the radius of the iterate set $\{x^k\}_{k=0,1,2,\dots}$ defined by*

$$R := \sup_{k=0,1,2,\dots} \|x^k - x^*\|$$

is bounded, that is, $R < +\infty$. Then the following inequality holds

$$T_{s+1}^0 \le T_s^m + \rho^{s/2} \cdot D_1 + \rho^s \cdot D_2, \tag{1}$$

where $D_1 = 2RL\big(1 + \frac{9L}{(3L-\mu)n}\big)\sqrt{\frac{\bar{L}^2}{\mu}}$, $D_2 = 4L\big(1 + \frac{9L}{2(3L-\mu)n}\big)\frac{\bar{L}^2}{\mu}$.

Proof. By the definition of T_s^k, we have

$$T_s^m = \frac{1}{n}\sum_{i=1}^n f_i(x_i^k) - f(\hat{x}_s^*) - \frac{1}{n}\sum_{i=1}^n \big\langle \nabla f_i(\hat{x}_s^*), x_i^k - \hat{x}_s^* \big\rangle + c_s\|x_s^m - \hat{x}_s^*\|^2, \tag{2}$$

and

$$T_{s+1}^0 = \frac{1}{n}\sum_{i=1}^n f_i(x_i^0) - f(\hat{x}_{s+1}^*) - \frac{1}{n}\sum_{i=1}^n \big\langle \nabla f_i(\hat{x}_{s+1}^*), x_i^0 - \hat{x}_{s+1}^* \big\rangle + c_{s+1}\|x_{s+1}^0 - \hat{x}_{s+1}^*\|^2. \tag{3}$$

Since $x_s^m = x_{s+1}^0$ and x_i^k in (2) is the same value as x_i^0 in (3), it holds that

$$\begin{aligned} T_{s+1}^0 \ - \ T_s^m = \ & f(\hat{x}_s^*) - f(\hat{x}_{s+1}^*) + \frac{1}{n}\sum_{i=1}^n \big\langle \nabla f_i(\hat{x}_s^*), x_i - \hat{x}_s^* \big\rangle \\ & -\frac{1}{n}\sum_{i=1}^n \big\langle \nabla f_i(\hat{x}_{s+1}^*), x_i - \hat{x}_{s+1}^* \big\rangle + c_{s+1}\|x_{s+1}^0 - \hat{x}_{s+1}^*\|^2 - c_s\|x_s^m - \hat{x}_s^*\|^2. \end{aligned} \tag{4}$$

We combine the two inner product terms on the right side:

$$\begin{aligned} & \big\langle \nabla f_i(\hat{x}_s^*), x_i - \hat{x}_s^* \big\rangle - \big\langle \nabla f_i(\hat{x}_{s+1}^*), x_i - \hat{x}_{s+1}^* \big\rangle \\ = \ & \big\langle \nabla f_i(\hat{x}_s^*) - \nabla f_i(\hat{x}_{s+1}^*) + \nabla f_i(\hat{x}_{s+1}^*), x_i - \hat{x}_s^* \big\rangle - \big\langle \nabla f_i(\hat{x}_{s+1}^*), x_i - \hat{x}_{s+1}^* \big\rangle \\ = \ & \big\langle \nabla f_i(\hat{x}_s^*) - \nabla f_i(\hat{x}_{s+1}^*), x_i - \hat{x}_s^* \big\rangle + \big\langle \nabla f_i(\hat{x}_{s+1}^*), \hat{x}_{s+1}^* - \hat{x}_s^* \big\rangle, \end{aligned} \tag{5}$$

where $\langle \nabla f_i(\hat{x}_s^*) - \nabla f_i(\hat{x}_{s+1}^*), x_i - \hat{x}_s^* \rangle$ can be bounded by

$$\begin{aligned}
&\langle \nabla f_i(\hat{x}_s^*) - \nabla f_i(\hat{x}_{s+1}^*), x_i - \hat{x}_s^* \rangle \\
\le\ & \|\nabla f_i(\hat{x}_s^*) - \nabla f_i(\hat{x}_{s+1}^*)\| \cdot \|x_i - \hat{x}_s^*\| \\
\le\ & \|\nabla f_i(\hat{x}_s^*) - \nabla f_i(\hat{x}_{s+1}^*)\| \cdot (\|x_i - x^*\| + \|\hat{x}_s^* - x^*\|) \\
\le\ & L\|\hat{x}_s^* - \hat{x}_{s+1}^*\| \cdot (R + \|\hat{x}_s^* - x^*\|),
\end{aligned} \tag{6}$$

the last inequality holds since f_i is L-smooth, and $\langle \nabla f_i(\hat{x}_{s+1}^*), \hat{x}_{s+1}^* - \hat{x}_s^* \rangle$ can be bounded by

$$\langle \nabla f_i(\hat{x}_{s+1}^*), \hat{x}_{s+1}^* - \hat{x}_s^* \rangle \le -f_i(\hat{x}_s^*) + f_i(\hat{x}_{s+1}^*) + \frac{L}{2}\|\hat{x}_s^* - \hat{x}_{s+1}^*\|^2, \tag{7}$$

which is also due to the property that f_i is L-smooth. Meanwhile, we have

$$\begin{aligned}
&c_{s+1}\|x_{s+1}^0 - \hat{x}_{s+1}^*\|^2 - c_s\|x_s^m - \hat{x}_s^*\|^2 \\
\le\ & c_s\|x_{s+1}^0 - \hat{x}_{s+1}^*\|^2 - c_s\|x_s^m - \hat{x}_s^*\|^2 \\
=\ & c_s\langle \hat{x}_s^* - \hat{x}_{s+1}^*, x_{s+1}^0 + x_s^m - \hat{x}_{s+1}^* - \hat{x}_s^* \rangle \\
\le\ & c_s\|\hat{x}_s^* - \hat{x}_{s+1}^*\| \cdot \|x_{s+1}^0 + x_s^m - \hat{x}_{s+1}^* - \hat{x}_s^*\| \\
\le\ & c_s\|\hat{x}_s^* - \hat{x}_{s+1}^*\| \cdot \big(\|x_{s+1}^0 - x^*\| + \|\hat{x}_{s+1}^* - x^*\| + \|x_s^m - x^*\| + \|\hat{x}_s^* - x^*\|\big) \\
\le\ & c_s\|\hat{x}_s^* - \hat{x}_{s+1}^*\| \cdot (2R + \|\hat{x}_{s+1}^* - x^*\| + \|\hat{x}_s^* - x^*\|),
\end{aligned} \tag{8}$$

the last inequality holds due to the assumption that $\|x^k - x^*\|_{k=0,1,2,\ldots}$ is not greater than R and $R < +\infty$. Plugging (5), (6), (7) and (8) into (4) yields

$$\begin{aligned}
T_{s+1}^0 - T_s^m\ \le\ & L\|\hat{x}_s^* - \hat{x}_{s+1}^*\| \cdot (R + \|\hat{x}_s^* - x^*\|) + \frac{L}{2}\|\hat{x}_s^* - \hat{x}_{s+1}^*\|^2 \\
&+ c_s\|\hat{x}_s^* - \hat{x}_{s+1}^*\| \cdot (2R + \|\hat{x}_{s+1}^* - x^*\| + \|\hat{x}_s^* - x^*\|).
\end{aligned} \tag{9}$$

We further bound $\|\hat{x}_s^* - x^*\|^2$ by

$$\|\hat{x}_s^* - x^*\|^2 \le \frac{2}{\mu}(F(\hat{x}_s^*) - F(x^*)) \le \frac{2}{\mu}(F(\hat{x}_s^*) - \hat{F}(\hat{x}_s^*)) \le \frac{\bar{L}^2}{\mu}\gamma_s, \tag{10}$$

where the three inequalities are due to the strong convexity of F, the fact that $F(x^*) \ge \hat{F}(x^*) \ge \hat{F}(\hat{x}_s^*)$, and Lemma 1 respectively. Bounding $\|\hat{x}_{s+1}^* - x^*\|^2$ in the same way as (10), we get

$$\|\hat{x}_{s+1}^* - x^*\|^2 \le \frac{\bar{L}^2}{\mu}\gamma_{s+1}, \tag{11}$$

Note that $\|\hat{x}_s^* - \hat{x}_{s+1}^*\| \le \|\hat{x}_s^* - x^*\| + \|\hat{x}_{s+1}^* - x^*\|$, $\|\hat{x}_s^* - \hat{x}_{s+1}^*\|^2 \le 2\|\hat{x}_s^* - x^*\|^2 + 2\|\hat{x}_{s+1}^* - x^*\|^2$ and $\gamma_{s+1} \le \gamma_s$. Plugging the above three inequalities, (10) and (11) into (9), we get

$$\begin{aligned}
T_{s+1}^0 - T_s^m \quad \le \quad & L \cdot 2\sqrt{\frac{\bar{L}^2}{\mu}\gamma_s} \cdot \big(R + \sqrt{\frac{\bar{L}^2}{\mu}\gamma_s}\big) + \frac{L}{2} \cdot 4\frac{\bar{L}^2}{\mu}\gamma_s \\
& + c_s \cdot 2\sqrt{\frac{\bar{L}^2}{\mu}\gamma_s} \cdot \big(2R + 2\sqrt{\frac{\bar{L}^2}{\mu}\gamma_s}\big) \\
=\ & \big(2LR + 4Rc_s\big)\sqrt{\frac{\bar{L}^2}{\mu}\gamma_s} + 4(L + c_s)\frac{\bar{L}^2}{\mu}\gamma_s \\
\le\ & 2RL\big(1 + \frac{9L}{(3L-\mu)n}\big)\sqrt{\frac{\bar{L}^2}{\mu}}\rho^{s/2} + 4L\big(1 + \frac{9L}{2(3L-\mu)n}\big)\frac{\bar{L}^2}{\mu}\rho^s.
\end{aligned}$$

$\square$

2 Proof of Theorem 2.

Theorem 2 (APA-SAGA)**.** *Under the same assumptions of Lemma 2, the following inequality holds*

$$\begin{aligned} &\mathbb{E}\|x_s - x^*\|^2 \\ \leq\ & \frac{4n}{3L}T_0^0 \cdot \theta^{s+1} + \frac{\bar{L}^2}{\mu}\frac{2}{3L}\rho^s + \theta\frac{\theta^s - \rho^{s/2}}{\theta - \rho^{1/2}} \cdot \frac{4n}{3L}D_1 + \theta\frac{\theta^s - \rho^s}{\theta - \rho} \cdot \frac{4n}{3L}D_2. \end{aligned} \tag{12}$$

Proof. By Young's inequality, it holds that

$$\|x_s - x^*\|^2 \leq 2\|x_s - \hat{x}_s^*\|^2 + 2\|\hat{x}_s^* - x^*\|^2, \tag{13}$$

we bound the two terms on the right side respectively. The first term can be bounded by

$$\|\hat{x}_s^* - x^*\|^2 \leq \frac{2}{\mu}(F(\hat{x}_s^*) - F(x^*)) \leq \frac{2}{\mu}(F(\hat{x}_s^*) - \hat{F}(\hat{x}_s^*)) \leq \frac{\bar{L}^2}{\mu}\gamma_s, \tag{14}$$

where the three inequalities are due to the strong convexity of F, the fact that $F(x^*) \geq \hat{F}(x^*) \geq \hat{F}(\hat{x}_s^*)$, and Lemma 1 respectively.

Meanwhile, according to Lemma 2, we have

$$\mathbb{E}T_s^m \leq \theta \cdot \mathbb{E}T_s^0 \leq \theta \cdot \mathbb{E}\big(T_{s-1}^m + \rho^{(s-1)/2}D_1 + \rho^{s-1}D_2\big).$$

Summing the above inequality over $0, 1, \dots, s$, we get

$$\begin{aligned} \mathbb{E}T_s^m \ &\leq\ \theta^s \cdot \mathbb{E}T_0^m + \big(\theta\rho^{s-1} + \theta^2\rho^{s-2} + \cdots + \theta^s\big) \cdot D_2 \\ &\quad + \big(\theta\rho^{(s-1)/2} + \theta^2\rho^{(s-2)/2} + \cdots + \theta^s\big) \cdot D_1 \\ &\leq\ \theta^{s+1} \cdot T_0^0 + D_2 \cdot \theta\frac{\theta^s - \rho^s}{\theta - \rho} + D_1 \cdot \theta\frac{\theta^s - \rho^{s/2}}{\theta - \rho^{1/2}}. \end{aligned} \tag{15}$$

Note that $c\|x_s - \hat{x}^*\| \leq T_s^m$ and $c \geq \frac{3L}{2n}$, we can further deduce from (15) that

$$\mathbb{E}\|x_s - \hat{x}_s^*\|^2 \leq \frac{2n}{3L}T_0^0 \cdot \theta^{s+1} + \theta\frac{\theta^s - \rho^{s/2}}{\theta - \rho^{1/2}} \cdot \frac{2n}{3L}D_1 + \theta\frac{\theta^s - \rho^s}{\theta - \rho} \cdot \frac{2n}{3L}D_2. \tag{16}$$

Plugging (14) and (16) into (13) with taking expectation on each term leads to the result and completes the proof. □