Supplementary Material

We will make use of the following result throughout this section.

Proposition S1 *Suppose f is L-smooth and m-strongly convex. Then for all x, y the following inequalities hold.*

$$f(x) - f(y) \geq \nabla f(y)^\mathsf{T}(x-y) + \frac{m}{2}\|x-y\|^2 \tag{S1}$$

$$f(y) - f(x) \geq \nabla f(y)^\mathsf{T}(y-x) - \frac{L}{2}\|y-x\|^2 \tag{S2}$$

Proof. These inequalities follow from the definitions of L-smoothness and m-strong convexity. ∎

A. Proof of Lemma 3

Applying (S1) with $(x, y) \mapsto (x_k, y_k)$, we obtain

$$f(x_k) - f(y_k) \geq \nabla f(y_k)^\mathsf{T}(x_k - y_k) + \frac{m}{2}\|x_k - y_k\|^2.$$

Applying (S2) with $(x, y) \mapsto (y_k - \alpha\nabla f(y_k), y_k)$, we obtain

$$f(y_k) - f(y_k - \alpha\nabla f(y_k)) \geq \frac{\alpha}{2}(2 - L\alpha)\|\nabla f(y_k)\|^2.$$

Summing these inequalities, we obtain:

$$f(x_k) - f(y_k - \alpha\nabla f(y_k)) \geq \nabla f(y_k)^\mathsf{T}(x_k - y_k) + \frac{m}{2}\|x_k - y_k\|^2 + \frac{\alpha}{2}(2 - L\alpha)\|\nabla f(y_k)\|^2. \tag{S3}$$

Substituting $x_{k+1} = y_k - \alpha\nabla f(y_k)$ in the left-hand side of (S3), we can rewrite it as

$$\frac{1}{2}\begin{bmatrix} x_k - y_k \\ \nabla f(y_k) \end{bmatrix}^\mathsf{T} \left(\begin{bmatrix} m & 1 \\ 1 & \alpha(2 - L\alpha) \end{bmatrix} \otimes I_p \right) \begin{bmatrix} x_k - y_k \\ \nabla f(y_k) \end{bmatrix} \leq f(x_k) - f(x_{k+1}). \tag{S4}$$

Substituting $y_k = (1+\beta)x_k - \beta x_{k-1}$ into (S4), we obtain

$$\frac{1}{2}\begin{bmatrix} x_k - x_\star \\ x_{k-1} - x_\star \\ \nabla f(y_k) \end{bmatrix}^\mathsf{T} \left(\begin{bmatrix} \beta^2 m & -\beta^2 m & -\beta \\ -\beta^2 m & \beta^2 m & \beta \\ -\beta & \beta & \alpha(2 - L\alpha) \end{bmatrix} \otimes I_p \right) \begin{bmatrix} x_k - x_\star \\ x_{k-1} - x_\star \\ \nabla f(y_k) \end{bmatrix} \leq f(x_k) - f(x_{k+1}),$$

which directly leads to the formulation of $\tilde{X}_1$ in Lemma 3. Similarly, we apply (S1) with $(x, y) \mapsto (x_\star, y_k)$ and obtain

$$\frac{1}{2}\begin{bmatrix} x_k - x_\star \\ x_{k-1} - x_\star \\ \nabla f(y_k) \end{bmatrix}^\mathsf{T} \left(\begin{bmatrix} (1+\beta)^2 m & -\beta(1+\beta)m & -(1+\beta) \\ -\beta(1+\beta)m & \beta^2 m & \beta \\ -(1+\beta) & \beta & \alpha(2 - L\alpha) \end{bmatrix} \otimes I_p \right) \begin{bmatrix} x_k - x_\star \\ x_{k-1} - x_\star \\ \nabla f(y_k) \end{bmatrix} \leq f(x_\star) - f(x_{k+1})$$

which directly leads to the formulation of $\tilde{X}_2$ in Lemma 3. The rest of the proof is straightforward. Actually, we can choose $\tilde{X} := \rho^2\tilde{X}_1 + (1-\rho^2)\tilde{X}_2$ and we directly obtain

$$\begin{bmatrix} x_k - x_\star \\ x_{k-1} - x_\star \\ \nabla f(y_k) \end{bmatrix}^\mathsf{T} \left(\tilde{X} \otimes I_p \right) \begin{bmatrix} x_k - x_\star \\ x_{k-1} - x_\star \\ \nabla f(y_k) \end{bmatrix} \leq -(f(x_{k+1}) - f(x_\star)) + \rho^2(f(x_k) - f(x_\star)).$$

Specifically, $\tilde{X}$ may be computed as

$$\tilde{X} = \frac{1}{2}\begin{bmatrix} (1+\beta)^2 m - (1+2\beta)m\rho^2 & (\rho^2 - 1 - \beta)\beta m & \rho^2 - 1 - \beta \\ (\rho^2 - 1 - \beta)\beta m & \beta^2 m & \beta \\ \rho^2 - 1 - \beta & \beta & \alpha(2 - L\alpha) \end{bmatrix}.$$

∎

B. Proof of Lemma 5

Applying (S2) with $(x, y) \mapsto (x_{k+1}, y_k)$, and making the substitutions $x_{k+1} = (1+\beta)x_k - \beta x_{k-1} - \alpha \nabla f(y_k)$ and $y_k = (1+\eta)x_k - \eta x_{k-1}$, we obtain:

$$
\begin{aligned}
f(y_k) - f(x_{k+1}) &\geq \nabla f(y_k)^\mathsf{T}(y_k - x_{k+1}) - \frac{L}{2}\|x_{k+1} - y_k\|^2 \\
&= \nabla f(y_k)^\mathsf{T}((\beta-\eta)(x_{k-1} - x_k) + \alpha \nabla f(y_k)) - \frac{L}{2}\|(\beta-\eta)(x_{k-1} - x_k) + \alpha \nabla f(y_k)\|^2 \\
&= \frac{1}{2}\begin{bmatrix} x_k - x_\star \\ x_{k-1} - x_\star \\ \nabla f(y_k) \end{bmatrix}^\mathsf{T} \left(\begin{bmatrix} -L(\beta-\eta)^2 & L(\beta-\eta)^2 & -(1-L\alpha)(\beta-\eta) \\ L(\beta-\eta)^2 & -L(\beta-\eta)^2 & (1-L\alpha)(\beta-\eta) \\ -(1-L\alpha)(\beta-\eta) & (1-L\alpha)(\beta-\eta) & \alpha(2-L\alpha) \end{bmatrix} \otimes I_p \right) \begin{bmatrix} x_k - x_\star \\ x_{k-1} - x_\star \\ \nabla f(y_k) \end{bmatrix}
\end{aligned} \tag{S5}
$$

Applying (S1) with $(x, y) \mapsto (x_k, y_k)$ and substituting $y_k = (1+\eta)x_k - \eta x_{k-1}$, we obtain:

$$
\begin{aligned}
f(x_k) - f(y_k) &\geq \nabla f(y_k)^\mathsf{T}(x_k - y_k) + \frac{m}{2}\|x_k - y_k\|^2 \\
&= \eta \nabla f(y_k)^\mathsf{T}(x_{k-1} - x_k) + \frac{m\eta^2}{2}\|x_{k-1} - x_k\|^2 \\
&= \frac{1}{2}\begin{bmatrix} x_k - x_\star \\ x_{k-1} - x_\star \\ \nabla f(y_k) \end{bmatrix}^\mathsf{T} \left(\begin{bmatrix} \eta^2 m & -\eta^2 m & -\eta \\ -\eta^2 m & \eta^2 m & \eta \\ -\eta & \eta & 0 \end{bmatrix} \otimes I_p \right) \begin{bmatrix} x_k - x_\star \\ x_{k-1} - x_\star \\ \nabla f(y_k) \end{bmatrix}
\end{aligned} \tag{S6}
$$

Applying (S1) with $(x, y) \mapsto (x_\star, y_k)$ and again substituting $y_k = (1+\eta)x_k - \eta x_{k-1}$, we obtain:

$$
\begin{aligned}
f(x_\star) - f(y_k) &\geq \nabla f(y_k)^\mathsf{T}(x_\star - y_k) + \frac{m}{2}\|x_\star - y_k\|^2 \\
&= -\nabla f(y_k)^\mathsf{T}((1+\eta)(x_k - x_\star) - \eta(x_{k-1} - x_\star)) + \frac{m}{2}\|(1+\eta)(x_k - x_\star) - \eta(x_{k-1} - x_\star)\|^2 \\
&= \frac{1}{2}\begin{bmatrix} x_k - x_\star \\ x_{k-1} - x_\star \\ \nabla f(y_k) \end{bmatrix}^\mathsf{T} \left(\begin{bmatrix} (1+\eta)^2 m & -\eta(1+\eta)m & -(1+\eta) \\ -\eta(1+\eta)m & \eta^2 m & \eta \\ -(1+\eta) & \eta & 0 \end{bmatrix} \otimes I_p \right) \begin{bmatrix} x_k - x_\star \\ x_{k-1} - x_\star \\ \nabla f(y_k) \end{bmatrix}
\end{aligned} \tag{S7}
$$

By adding (S5)–(S7) with the definitions of $\tilde{X}_1$, $\tilde{X}_2$, and $\tilde{X}_3$ in Lemma 5, we obtain:

$$
\begin{bmatrix} x_k - x_\star \\ x_{k-1} - x_\star \\ \nabla f(y_k) \end{bmatrix}^\mathsf{T} \left((\tilde{X}_1 + \tilde{X}_2) \otimes I_p \right) \begin{bmatrix} x_k - x_\star \\ x_{k-1} - x_\star \\ \nabla f(y_k) \end{bmatrix} \leq f(x_k) - f(x_{k+1})
$$

$$
\begin{bmatrix} x_k - x_\star \\ x_{k-1} - x_\star \\ \nabla f(y_k) \end{bmatrix}^\mathsf{T} \left((\tilde{X}_1 + \tilde{X}_3) \otimes I_p \right) \begin{bmatrix} x_k - x_\star \\ x_{k-1} - x_\star \\ \nabla f(y_k) \end{bmatrix} \leq f(x_\star) - f(x_{k+1})
$$

The rest of the proof follows by substituting above expressions into the weighted sum with ρ^2. ■

C. Proof of Lemma 8

Since f is L-smooth and convex, we can use the same proof technique as in Lemma 3 while setting $m = 0$ and $\alpha = \frac{1}{L}$. We can thus obtain the following inequalities that parallel (S4).

$$
\frac{1}{2}\begin{bmatrix} y_k - x_k \\ \nabla f(y_k) \end{bmatrix}^\mathsf{T} \left(\begin{bmatrix} 0 & 1 \\ 1 & -\frac{1}{L} \end{bmatrix} \otimes I_p \right) \begin{bmatrix} y_k - x_k \\ \nabla f(y_k) \end{bmatrix} \geq f(x_{k+1}) - f(x_k)
$$

$$
\frac{1}{2}\begin{bmatrix} y_k - x_\star \\ \nabla f(y_k) \end{bmatrix}^\mathsf{T} \left(\begin{bmatrix} 0 & 1 \\ 1 & -\frac{1}{L} \end{bmatrix} \otimes I_p \right) \begin{bmatrix} y_k - x_\star \\ \nabla f(y_k) \end{bmatrix} \geq f(x_{k+1}) - f(x_\star)
$$

The conclusion of Lemma 8 follows once we substitute $y_k = (1 - \beta_k)x_k + \beta_k x_{k-1}$. ■