Appendix

A Proof for Lemma 3

1. The statement follows readily from the following facts. Smoothness implies that if $\alpha^+ \leq \frac{\delta}{L_f}$, Algorithm 2 terminates. Therefore, it must be $\alpha^+ \geq \frac{\delta}{L_f}$.

2. The lower bound in (8) follows readily from the established inequality $\alpha^+ \geq \frac{\delta}{L_f}$ and $\alpha^+ \leq \gamma\alpha$. Furthermore, it follows from the strong convexity of f that

$$f(x^+) \geq f(x) + \left\langle \nabla f(x), x^+ - x \right\rangle + \frac{\mu}{2}\|x^+ - x\|^2.$$

Therefore, α^+ cannot exceed $\frac{\delta}{\mu}$, which together with $\alpha^+ \leq \gamma\alpha$ proves the upper bound in (8).

3.Let us introduce the function $f_d(x) = f(x) + \langle d, x\rangle$. Notice that f_d inherits the same smoothness and (strong) convexity property of f. The termination condition in Algorithm 2 can be rewritten in terms of $f_d(x)$ as following:

$$f_d(x^+) \leq f_d(x) + \left\langle \nabla f_d(x), x^+ - x \right\rangle + \frac{\delta}{2\alpha}\|x^+ - x\|^2.$$

Using $x^+ - x = d$, the condition above reads

$$f_d(x - \alpha^+ d) \leq f_d(x) - \alpha^+ \left(1 - \frac{\delta}{2}\right)\|d\|^2. \tag{13}$$

Because of convexity of f_d, if (13) holds for some α^+, then it holds for any $\alpha \in [0, \alpha^+]$. □

B Proof of Theorem 4

We establish the proof under the weaker requirements on f_i than what stated in Assumption 1 and postulated in the theorem, namely each f_i is assumed *locally* smooth and strongly convex.

Our first result determines the relationship between the values of the merit function (11) in two consecutive iterations. We have the following.

Lemma 7. *The update* $(\mathbf{D}^k, \mathbf{X}^k) \to (\mathbf{D}^{k+1}, \mathbf{X}^{k+1})$ *satisfies*

$$\begin{aligned} V^{k+1} = &\left\|\mathbf{X}^k - \mathbf{X}^\star\right\|^2 + (\alpha^k)^2\|\mathbf{D}^k - \mathbf{D}^\star\|_M^2 \\ &- \|\mathbf{X}^k - \mathbf{X}^{k+1}\|^2 - (\alpha^k)^2\|\mathbf{D}^k - \mathbf{D}^{k+1}\|_M^2 \\ &+ 2\alpha^k \left\langle \nabla F(\mathbf{X}^{k+1/2}) - \nabla F(\mathbf{X}^\star), \mathbf{X}^\star - \mathbf{X}^{k+1} \right\rangle. \end{aligned} \tag{14}$$

Proof. See Appendix C.1. □

We proceed to bound the inner-product above in terms the X-sequence. Invoking the properties of the backtracking (Lemma 8) and local strong convexity, we obtain the following.

Lemma 8. *The following holds:*

$$\left\langle \nabla F(\mathbf{X}^{k+1/2}) - \nabla F(\mathbf{X}^\star), \mathbf{X}^\star - \mathbf{X}^{k+1} \right\rangle \leq \frac{\delta}{2\alpha^k}\|\mathbf{X}^{k+1} - \mathbf{X}^{k+1/2}\|^2 - \frac{\mu^k}{2}\|\mathbf{X}_k^{k+1/2} - \mathbf{X}^\star\|^2,$$

where μ^k is the local strong convexity constant of f_i along the segment $[\mathbf{X}^{k+1/2}, \mathbf{X}^\star]$.

Proof. See Appendix C.2. □

Combining Lemma 7 and Lemma 8, we obtain the following.

Lemma 9. *For any $c \leq 1/2$ and $\{\gamma^k \geq 1\}$, it holds:*

$$\begin{aligned} V^{k+1} \leq & \left\|\mathbf{X}^k - \mathbf{X}^\star\right\|^2 + (\alpha^k)^2 \|\mathbf{D}^k - \mathbf{D}^\star\|_M^2 \\ & - \mu^k \alpha^k \|\mathbf{X}_k^{k+1/2} - \mathbf{X}^\star\|^2 \\ & - \left\langle \mathbf{X}^k, c(I - \widetilde{W})\mathbf{X}^k \right\rangle - (\alpha^k)^2 \left\| c(I - \widetilde{W}) \left(\nabla F(\mathbf{X}^{k+1/2}) + \mathbf{D}^k \right) \right\|_M^2 . \end{aligned} \tag{15}$$

Proof. See Appendix C.3. $\square$

To achieve contraction in (15) we split the aiding term $-\mu^k\alpha^k\|\mathbf{X}_k^{k+1/2} - \mathbf{X}^\star\|^2 = -\frac{\mu^k\alpha^k}{2}\|\mathbf{X}_k^{k+1/2} - \mathbf{X}^\star\|^2 - \frac{\mu^k\alpha^k}{2}\|\mathbf{X}_k^{k+1/2} - \mathbf{X}^\star\|^2$ and use each to control primal and dual errors, as detailed next.

Define $A := c(I - \widetilde{W})$ for notational convenience. Using the definitions of r^k and g^k as given in (10) and (9), respectively, the following lemma allows one to control the last two terms in (15) via $\frac{\mu^k\alpha^k}{2}\|\mathbf{X}_k^{k+1/2} - \mathbf{X}^\star\|^2$.

Lemma 10. *For any $c < 1/2$ and $\{\gamma^k \geq 1\}$, it holds:*

$$\begin{aligned} & -\frac{\mu^k\alpha^k}{2}\|\mathbf{X}^{k+1/2} - \mathbf{X}^\star\|^2 - \langle \mathbf{X}^k, A\mathbf{X}^k \rangle - (\alpha^k)^2 \left\| A \left(\nabla F(\mathbf{X}^{k+1/2}) + \mathbf{D}^k \right) \right\|_M^2 \\ \leq & -\max\left((r^k)^2, \frac{\mu^k\alpha^k (g^k[1 - r^k]_+)^2}{2} \right) c^2 (1 - \lambda_2(\widetilde{W}))^2 (\alpha^k)^2 \|(\mathbf{D}^k - \mathbf{D}^\star)\|_M^2 . \end{aligned} \tag{16}$$

Proof. See Appendix C.4. $\square$

It remains to control $\left\|\mathbf{X}^k - \mathbf{X}^\star\right\|^2$ using $\frac{\mu^k\alpha^k}{2}\|\mathbf{X}_k^{k+1/2} - \mathbf{X}^\star\|^2$. Noting

$$\mathbf{X}^{k+1/2} = (I - A)\mathbf{X}^k \quad \text{and} \quad \mathbf{X}^\star = (I - A)\mathbf{X}^\star,$$

the term $\|\mathbf{X}^{k+1/2} - \mathbf{X}^\star\|^2$ can be lower bounded as

$$\|\mathbf{X}^{k+1/2} - \mathbf{X}^\star\|^2 \geq (1 - c(1 - \lambda_m(\widetilde{W})))^2 \|\mathbf{X}^k - \mathbf{X}^\star\|^2 . \tag{17}$$

Using in (9) the bounds derived in Lemma 10 and (17) yields

$$\begin{aligned} & V^{k+1} \\ \leq. & \left(1 - \mu^k\alpha^k \frac{(1 - c(1 - \lambda_n(\widetilde{W})))^2}{2} \right) \left\|\mathbf{X}^k - \mathbf{X}^\star\right\|_{I - c(I - \widetilde{W})}^2 \\ & + \left(1 - \max\left((r^k)^2, \frac{\mu^k\alpha^k (g^k[1 - r^k]_+)^2}{2} \right) c^2 (1 - \lambda_2(\widetilde{W}))^2 \right) (\alpha^k)^2 \|\mathbf{D}^k - \mathbf{D}^\star\|_M^2 \\ \leq & \left(1 - \min\left[\mu^k\alpha^k \frac{(1 - c(1 - \lambda_n(\widetilde{W})))^2}{2}, \max\left((r^k)^2, \frac{\mu^k\alpha^k (g^k[1 - r^k]_+)^2}{2} \right) c^2 (1 - \lambda_2(\widetilde{W}))^2 \right] \right) \\ & \times \max\left[1, \left(\frac{\alpha^k}{\alpha_{k-1}} \right)^2 \right] \times V^k . \end{aligned} \tag{18}$$

This proves (12).

Since f_i's are assumed to be only locally smooth and strongly convex, it remains to show that the sequence generated by the algorithm is bounded. This is proved below under the condition $r^k > r$, for some r. Notice that if the f_i are globally smooth and strongly convex, boundedness of the iterates is granted with no extra conditions.

Lemma 11. *In the setting above, further assume that $r_k \geq r$, for some $r > 0$. Then, sequence $\{\mathbf{X}^k, \mathbf{D}^k\}$ generated by Algorithm 1 is bounded.*

Proof. See Appendix D. $\square$

This completes the proof of 4. $\square$

C Proof of the Intermediate Results in Appendix B

C.1 Proof of Lemma 7

From the update of $\mathbf{X}^{k+1}$ and the facts $\mathbf{D}^{k+1} - \mathbf{D}^\star \in \mathtt{span}(I - \widetilde{W})$ and $\mathbf{X}^\star \in \mathtt{null}(I - \widetilde{W})$, we have:

$$\begin{aligned}
&\left\langle \mathbf{D}^{k+1} - \mathbf{D}^\star, \mathbf{X}^{k+1} - \mathbf{X}^\star \right\rangle \\
=&\left\langle \mathbf{D}^{k+1} - \mathbf{D}^\star, \mathbf{X}^{k} - \alpha^k \nabla F(\mathbf{X}^{k+1/2}) - \alpha^k \mathbf{D}^{k+1} - \mathbf{X}^\star \right\rangle \\
=&\left\langle \mathbf{D}^{k+1} - \mathbf{D}^\star, \alpha^k c^{-1} \left(I - \widetilde{W}\right)^\dagger (\mathbf{D}^{k+1} - \mathbf{D}^k) + \alpha^k \mathbf{D}^k - \alpha^k \mathbf{D}^{k+1} \right\rangle \\
=&\alpha^k \left\langle \mathbf{D}^{k+1} - \mathbf{D}^\star, \mathbf{D}^{k+1} - \mathbf{D}^k \right\rangle_M .
\end{aligned} \tag{19}$$

Using the equality above, we can write

$$\begin{aligned}
&\alpha^k \left\langle \mathbf{X}^k - \mathbf{X}^\star, \nabla F(\mathbf{X}^{k+1/2}) - \nabla F(\mathbf{X}^\star) \right\rangle \\
=&\alpha^k \left\langle \mathbf{X}^k - \mathbf{X}^\star, \frac{1}{\alpha^k}\left(\mathbf{X}^k - \mathbf{X}^{k+1}\right) - \mathbf{D}^{k+1} + \mathbf{D}^\star \right\rangle \\
=&\left\langle \mathbf{X}^k - \mathbf{X}^\star, \mathbf{X}^k - \mathbf{X}^{k+1} \right\rangle - \alpha^k \left\langle \mathbf{D}^{k+1} - \mathbf{D}^\star, \mathbf{X}^k - \mathbf{X}^\star \right\rangle \\
=&\left\langle \mathbf{X}^k - \mathbf{X}^\star, \mathbf{X}^k - \mathbf{X}^{k+1} \right\rangle + \alpha^k \left\langle \mathbf{D}^{k+1} - \mathbf{D}^\star, \mathbf{X}^{k+1} - \mathbf{X}^k \right\rangle - \alpha^k \left\langle \mathbf{D}^{k+1} - \mathbf{D}^\star, \mathbf{X}^{k+1} - \mathbf{X}^\star \right\rangle \\
=&\left\langle \mathbf{X}^k - \mathbf{X}^\star, \mathbf{X}^k - \mathbf{X}^{k+1} \right\rangle + \alpha^k \left\langle \mathbf{D}^{k+1} - \mathbf{D}^\star, \mathbf{X}^{k+1} - \mathbf{X}^k \right\rangle - (\alpha^k)^2 \left\langle \mathbf{D}^{k+1} - \mathbf{D}^\star, \mathbf{D}^{k+1} - \mathbf{D}^k \right\rangle_M \\
=&\left\langle (\mathbf{X}^k - \alpha^k \mathbf{D}^{k+1}) - \mathbf{X}^\star + \alpha^k \mathbf{D}^\star, \mathbf{X}^k - \mathbf{X}^{k+1} \right\rangle - (\alpha^k)^2 \left\langle \mathbf{D}^{k+1} - \mathbf{D}^\star, \mathbf{D}^{k+1} - \mathbf{D}^k \right\rangle_M \\
=&\left\langle \mathbf{X}^{k+1} - \mathbf{X}^\star + \alpha^k \nabla F(\mathbf{X}^{k+1/2}) - \alpha^k \nabla F(\mathbf{X}^\star), \mathbf{X}^k - \mathbf{X}^{k+1} \right\rangle - (\alpha^k)^2 \left\langle \mathbf{D}^{k+1} - \mathbf{D}^\star, \mathbf{D}^{k+1} - \mathbf{D}^k \right\rangle_M \\
=&\left\langle \mathbf{X}^{k+1} - \mathbf{X}^\star, \mathbf{X}^k - \mathbf{X}^{k+1} \right\rangle + \alpha^k \left\langle \nabla F(\mathbf{X}^{k+1/2}) - \nabla F(\mathbf{X}^\star), \mathbf{X}^k - \mathbf{X}^{k+1} \right\rangle \\
&\quad - (\alpha^k)^2 \left\langle \mathbf{D}^{k+1} - \mathbf{D}^\star, \mathbf{D}^{k+1} - \mathbf{D}^k \right\rangle_M ,
\end{aligned}$$

where the first and fifth equations follow from the update of $\mathbf{X}^{k+1}$ while in the third equation we used (19).

Finally, rearranging the terms above, yields

$$\begin{aligned}
&- 2\alpha^k \left\langle \nabla F(\mathbf{X}^{k+1/2}) - \nabla F(\mathbf{X}^\star), \mathbf{X}^\star - \mathbf{X}^{k+1} \right\rangle \\
&= 2 \left\langle \mathbf{X}^{k+1} - \mathbf{X}^\star, \mathbf{X}^k - \mathbf{X}^{k+1} \right\rangle - 2(\alpha^k)^2 \left\langle \mathbf{D}^{k+1} - \mathbf{D}^\star, \mathbf{D}^{k+1} - \mathbf{D}^k \right\rangle_M .
\end{aligned} \tag{20}$$

The final result (14) follows from (20) and $2\langle \mathbf{a}, \mathbf{b} \rangle = \|\mathbf{a} + \mathbf{b}\|^2 - \|a\|^2 - \|b\|^2$. $\square$

C.2 Proof for Lemma 8

Define

$$D(\mathbf{X}) = F(\mathbf{X}) + \langle \mathbf{D}^\star, \mathbf{X} \rangle, \tag{21}$$

where $(\mathbf{D}^\star, \mathbf{X}^\star)$ is a fixed-point of Algorithm 1, Using the definition of $D(\mathbf{X})$ and the fact that $\nabla F(\mathbf{X}^\star)+\mathbf{D}^\star=0$, $\langle \nabla F(\mathbf{X}^{k+1/2})-\nabla F(\mathbf{X}^\star), \mathbf{X}^\star - \mathbf{X}^{k+1}\rangle$ can be bounded as

$$
\begin{aligned}
&\left\langle \nabla F(\mathbf{X}^{k+1/2})-\nabla F(\mathbf{X}^\star), \mathbf{X}^\star-\mathbf{X}^{k+1}\right\rangle = \left\langle \nabla D(\mathbf{X}^{k+1/2}), \mathbf{X}^\star-\mathbf{X}^{k+1}\right\rangle\\
=&\left\langle \nabla D(\mathbf{X}^{k+1/2}), \mathbf{X}^\star-\mathbf{X}^{k+1/2}\right\rangle - \left\langle \nabla D(\mathbf{X}^{k+1/2}), \mathbf{X}^{k+1}-\mathbf{X}^{k+1/2}\right\rangle\\
\le& D(\mathbf{X}^\star)-D(\mathbf{X}^{k+1/2}) - \left\langle \nabla D(\mathbf{X}^{k+1/2}), \mathbf{X}^{k+1}-\mathbf{X}^{k+1/2}\right\rangle - \frac{\mu^k}{2}\|\mathbf{X}_k^{k+1/2}-\mathbf{X}^\star\|^2\\
\le& D(\mathbf{X}^{k+1})-D(\mathbf{X}^{k+1/2}) - \left\langle \nabla D(\mathbf{X}^{k+1/2}), \mathbf{X}^{k+1}-\mathbf{X}^{k+1/2}\right\rangle - \frac{\mu^k}{2}\|\mathbf{X}_k^{k+1/2}-\mathbf{X}^\star\|^2\\
=& D(\mathbf{X}^{k+1}) - \left(D(\mathbf{X}^{k+1/2}) + \left\langle \nabla D(\mathbf{X}^{k+1/2}), \mathbf{X}^{k+1}-\mathbf{X}^{k+1/2}\right\rangle + \frac{1}{2\alpha^k}\|\mathbf{X}^{k+1}-\mathbf{X}^{k+1/2}\|^2\right)\\
&\quad + \frac{\delta}{2\alpha^k}\|\mathbf{X}^{k+1}-\mathbf{X}^{k+1/2}\|^2 - \frac{\mu}{2}\|\mathbf{X}_k^{k+1/2}-\mathbf{X}^\star\|^2\\
=& F(\mathbf{X}^{k+1}) - \left(F(\mathbf{X}^{k+1/2}) + \left\langle \nabla F(\mathbf{X}^{k+1/2}), \mathbf{X}^{k+1}-\mathbf{X}^{k+1/2}\right\rangle + \frac{\delta}{2\alpha^k}\|\mathbf{X}^{k+1}-\mathbf{X}^{k+1/2}\|^2\right)\\
&\quad + \frac{\delta}{2\alpha^k}\|\mathbf{X}^{k+1}-\mathbf{X}^{k+1/2}\|^2 - \frac{\mu^k}{2}\|\mathbf{X}_k^{k+1/2}-\mathbf{X}^\star\|^2\\
\le& \frac{\delta}{2\alpha^k}\|\mathbf{X}^{k+1}-\mathbf{X}^{k+1/2}\|^2 - \frac{\mu^k}{2}\|\mathbf{X}_k^{k+1/2}-\mathbf{X}^\star\|^2,
\end{aligned}
$$

where the first inequality follow from the the strong convexity of $D(\mathbf{X})$ on the segment $[\mathbf{X}^{k+1/2}, \mathbf{X}^\star]$; and in the last inequality we used the algorithm update. □

C.3 Proof for Lemma 9

Uniting Lemma 7 and Lemma 8, we can write

$$
\begin{aligned}
V^{k+1} \le & \left\|\mathbf{X}^k-\mathbf{X}^\star\right\|^2 + (\alpha^k)^2\|\mathbf{D}^k-\mathbf{D}^\star\|_M^2\\
&- \mu^k\alpha^k\|\mathbf{X}_k^{k+1/2}-\mathbf{X}^\star\|^2\\
&- (1-\delta)\left(\|\mathbf{X}^k-\mathbf{X}^{k+1}\|^2 + (\alpha^k)^2\|\mathbf{D}^k-\mathbf{D}^{k+1}\|_M^2\right)\\
&+ \delta\left(\|\mathbf{X}^{k+1}-\mathbf{X}^{k+1/2}\|^2 - \|\mathbf{X}^k-\mathbf{X}^{k+1}\|^2 - (\alpha^k)^2\|\mathbf{D}^k-\mathbf{D}^{k+1}\|_M^2\right).
\end{aligned}
\tag{22}
$$

Consider the first and second term:

$$
\begin{aligned}
&\|\mathbf{X}^{k+1}-\mathbf{X}^{k+1/2}\|^2 - \|\mathbf{X}^k-\mathbf{X}^{k+1}\|^2\\
=&(\alpha^k)^2\left(\|\mathbf{D}^{k+1/2}\|^2 - \|\mathbf{D}^{k+1/2} + \frac{c}{\alpha^k}(I-\widetilde{W})\mathbf{X}^k\|^2\right)\\
=&-\|c(I-\widetilde{W})\mathbf{X}^k\|^2 - \alpha^k\left\langle \nabla\mathbf{D}^{k+1/2}, c(I-\widetilde{W})\mathbf{X}^k\right\rangle\\
=&-\|c(I-\widetilde{W})\mathbf{X}^k\|^2 - \alpha^k\left\langle \nabla\mathbf{D}^{k+1/2}, c(I-\widetilde{W})\mathbf{X}^k\right\rangle\\
=&-\|c(I-\widetilde{W})\mathbf{X}^k\|^2 + \alpha^k\left\langle \left(c(I-\widetilde{W})-I\right)\left(\nabla F(\mathbf{X}^{k+1/2})+\mathbf{D}^k\right), c(I-\widetilde{W})\mathbf{X}^k\right\rangle.
\end{aligned}
\tag{23}
$$

Proceeding with $\|\mathbf{D}^k - \mathbf{D}^{k+1}\|_M^2$, we have

$$
\begin{aligned}
&-(\alpha^k)^2\|\mathbf{D}^k - \mathbf{D}^{k+1}\|_M^2 \\
=&-(\alpha^k)^2\| - c(I-\widetilde{W})\left(\nabla F(\mathbf{X}^{k+1/2}) + \mathbf{D}^k\right) + \frac{c}{\alpha^k}(I-\widetilde{W})\mathbf{X}^k\|_M^2 \\
=&-\|c(I-\widetilde{W})\mathbf{X}^k\|_M^2 \\
&+\alpha^k\langle Mc(I-\widetilde{W})\left(\nabla F(\mathbf{X}^{k+1/2}) + \mathbf{D}^k\right), c(I-\widetilde{W})\mathbf{X}^k\rangle \\
&-(\alpha^k)^2\|c(I-\widetilde{W})\left(\nabla\nabla F(\mathbf{X}^{k+1/2}) + \mathbf{D}^k\right)\|_M^2 \\
=&-\left\|c(I-\widetilde{W})\mathbf{X}^k\right\|_M^2 \\
&+\alpha^k\left\langle\left(I-c(I-\widetilde{W})\right)\left(\nabla F(\mathbf{X}^{k+1/2}) + \mathbf{D}^k\right), c(I-\widetilde{W})\mathbf{X}^k\right\rangle \\
&-(\alpha^k)^2\left\|c(I-\widetilde{W})\left(\nabla F(\mathbf{X}^{k+1/2}) + \mathbf{D}^k\right)\right\|_M^2 \\
=&\left\|c(I-\widetilde{W})\mathbf{X}^k\right\|^2 - c\left\langle\mathbf{X}^k, \left(I-\widetilde{W}\right)\mathbf{X}^k\right\rangle \\
&+\alpha^k\left\langle\left(I-c(I-\widetilde{W})\right)\left(F\nabla(\mathbf{X}^{k+1/2}) + \mathbf{D}^k\right), c(I-\widetilde{W})\mathbf{X}^k\right\rangle \\
&-(\alpha^k)^2\left\|c(I-\widetilde{W})\left(\nabla F(\mathbf{X}^{k+1/2}) + \mathbf{D}^k\right)\right\|_M^2,
\end{aligned}
\tag{24}
$$

where last two equalities follow from the definition of $M = c^{-1}(I-\widetilde{W})^\dagger - I$.

Summing (23) and (24), we obtain

$$
\begin{aligned}
&\|\mathbf{X}^{k+1} - \mathbf{X}^{k+1/2}\|^2 - \|\mathbf{X}^k - \mathbf{X}^{k+1}\|^2 - (\alpha^k)^2\|\mathbf{D}^k - \mathbf{D}^{k+1}\|_M^2 \\
=&-c\left\langle\mathbf{X}^k, \left(I-\widetilde{W}\right)\mathbf{X}^k\right\rangle - (\alpha^k)^2\left\|c(I-\widetilde{W})\left(\nabla F(\mathbf{X}^{k+1/2}) + \mathbf{D}^k\right)\right\|_M^2.
\end{aligned}
$$

The proof follows readily from the above equality, (22), and setting $\delta = 1$. $\square$

C.4 Proof for Lemma 10

To obtain decrease for dual variable, we should consider two estimates.

1. Firstly, using the definition of r^k, we can bound the last two terms in (15) as

$$
\left\langle\mathbf{X}^k, A\mathbf{X}^k\right\rangle + (\alpha^k)^2\left\|A\left(\nabla F(\mathbf{X}^{k+1/2}) + \mathbf{D}^k\right)\right\|_M^2 \geq (\alpha^k)^2\,(r^k)^2\,\|A(\mathbf{D}^k - \mathbf{D}^\star)\|_M. \tag{25}
$$

2. Using again the definition of r^k and the reverse triangle inequality we have

$$
\begin{aligned}
r^k\|A(\mathbf{D}^k - \mathbf{D}^\star)\|_M &\geq \left\|A\left(\nabla F(\mathbf{X}^{k+1/2}) + \mathbf{D}^k\right)\right\|_M \\
&\geq \|A(\mathbf{D}^k - \mathbf{D}^\star)\|_M - \left\|A\left(\nabla F(\mathbf{X}^{k+1/2}) - F(\mathbf{X}^\star)\right)\right\|_M.
\end{aligned}
$$

   Therefore,

$$
\begin{aligned}
\left\|A\left(\nabla F(\mathbf{X}^{k+1/2}) - F(\mathbf{X}^\star)\right)\right\|_M &\geq [1-r_k]_+\|A(\mathbf{D}^k - \mathbf{D}^\star)\|_M \\
&\geq ([1-r_k]_+)\,c(1-\lambda_2(\widetilde{W})\,\|\mathbf{D}^k - \mathbf{D}^\star\|_M.
\end{aligned}
$$

   Using the definition of g_k yields

$$
\|\mathbf{X}^{k+1/2} - \mathbf{X}^\star\|^2 \geq (g^k[1-r^k]_+)^2(\alpha^k)^2\|A(\mathbf{D}^k - \mathbf{D}^\star)\|_M^2. \tag{26}
$$

The desired expression (16) follows combining (25) and (26).

D Proof of Lemma 11

According to Lemma 9, we have the following inequality:

$$V^{k+1} \leq \left\|\mathbf{X}^k - \mathbf{X}^\star\right\|^2 + (\alpha^k)^2 \|\mathbf{D}^k - \mathbf{D}^\star\|_M^2 - \left\langle \mathbf{X}^k, c(I - \widetilde{W})\mathbf{X}^k \right\rangle - (\alpha^k)^2 \left\|c(I - \widetilde{W}) \left(\nabla F(\mathbf{X}^{k+1/2}) + \mathbf{D}^k\right)\right\|_M^2. \tag{27}$$

Using definition of r^k, (27) can be written in the following form:

$$V^{k+1} \leq \left\|\mathbf{X}^k - \mathbf{X}^\star\right\|^2 + (\alpha^k)^2 (1 - r_k c^2 (1 - \lambda_2(\widetilde{W}))^2) \|\mathbf{D}^k - \mathbf{D}^\star\|_M^2.$$

Further, using $r_k \geq r$ and the definition of γ_k, we have

$$V^{k+1} \leq \max\left(1, \left(\frac{k+\beta_1}{k+1}\right)^{2\beta_2} (1 - rc^2(1 - \lambda_2(\widetilde{W}))^2)\right) V^k.$$

Starting from some iteration k^*, telescoping the above inequality to $k0$, yields

$$V^{k+1} \leq \prod_{j=0}^{k^*} \left(\left(\frac{k+\beta_1}{k+1}\right)^{2\beta_2} (1 - rc^2(1 - \lambda_2(\widetilde{W}))^2)\right) V^0 = R < \infty.$$

Using the definition of V^{k+1}, we have that $\|\mathbf{X}^{k+1} - \mathbf{X}^*\| \leq R$. It means that α^k is bounded below by the inverse local Lipschitz constant. So, $\|\mathbf{D}^{k+1} - \mathbf{D}^*\|$ is bounded too.

This completes the proof. □

E Proof for Corollary 4.1

Let us consider the value ρ_k from Theorem 4:

$$\rho^k := \min\left(\mu\,\alpha^k \frac{(1 - c(1 - \lambda_n(\widetilde{W})))^2}{2}, \max\left((r^k)^2, \mu\,\alpha^k \frac{(g^k\,[1 - r^k]_+)^2}{2}\right) c^2 (1 - \lambda_2(\widetilde{W}))^2\right).$$

If we prove that $\rho_k \geq \rho > 0$ then we can obtain an estimate for number of iterations to approach quality ε. Using the result of Theorem 4 and condition for γ_k, we have the following inequality:

$$(1-\rho)^k \prod_{j=0}^{k} \left(\frac{k+\beta_1}{k+1}\right)^{\beta_2} V^0 \leq \varepsilon$$

Note, for any β_1, β_2 and enough big k we have that $(1-\rho)^k \prod_{j=0}^{k} \left(\frac{k+\beta_1}{k+1}\right)^{\beta_2} V^0 \leq (1-\rho/2)^k V^0 \leq \exp(-k\rho/2) V^0$. It means, that to approach quality ε it is enough to perform

$$N = O(\frac{1}{\rho} \log(V^0/\varepsilon) \tag{28}$$

steps of Algorithm 1.

Furher, we consider different cases from Corollary 4.1 to obtain estimate for this value.

Firstly, note that according to Lemma 3 we have that $\alpha^k \geq 1/(2L)$.

1. If $r^k \geq \frac{1}{2}$ than we have that $\max\left((r^k)^2, \mu\,\alpha^k \frac{(g^k\,[1-r^k]_+)^2}{2}\right) \geq \frac{1}{4}$ and

$$\frac{1}{\rho} \leq O\left(\max\left(\frac{\kappa}{(1 - c(1 - \lambda_m(\widetilde{W})))^2}, \frac{1}{c^2(1 - \lambda_2(\widetilde{W}))^2}\right)\right)$$

2. If $r^k \geq (1/4)\sqrt{\kappa}$, then we have

$$\frac{1}{\rho} \leq O\left(\frac{\kappa}{\min\left(c(1 - \lambda_2(\widetilde{W})), (1 - c(1 - \lambda_m(\widetilde{W}))\right)^2}\right).$$

On the other hand, if $g^k \geq \frac{1}{2}$ we have the same result because $\mu\alpha^k \geq (1/2)\kappa$.

3. Finally, note that

$$\mu\alpha^k (g^k)^2 \geq \frac{\mu}{\alpha_k}\frac{1}{L^2} \geq \kappa^2,$$

where the first inequality holds because of smoothness and definition of g^k. The second because of upper bound on step size in Lemma 3. It immediately gives us

$$\frac{1}{\rho} \leq O\left(\left(\frac{\kappa}{\min\left(c, (1-c(1-\lambda_m(\widetilde{W})))\right)\cdot(1-\lambda_2(\widetilde{W}))}\right)^2\right)$$

These three cases and estimation (28) complete the proof.

F Proof of Theorem 5

Similarly to the proof of Theorem 4, we prove Theorem 5 under the weaker assumption that each f_i is locally smooth. Throughout the proof, we denote by L the smooth contant of f_i's over the convex hull of the set $\{\mathbf{X}^j\}_{j=0}^{k+1}$.

Define

$$q_k = \max\left(1, \frac{\alpha_k^2}{\alpha_{k-1}^2}\right).$$

According to Lemma 9, under $\mu = 0$, we have

$$V^{k+1} \leq V^k - (1-\delta)\left(\|\mathbf{X}^k - \mathbf{X}^{k+1}\|^2 + \alpha_k^2\|\mathbf{D}^k - \mathbf{D}^{k+1}\|_M^2\right).$$

Applying recursively the above inequality, yields

$$V^{k+1} \leq \left(\prod_{j=0}^{k} q_j\right) V^0 - (1-\delta)\sum_{j=0}^{k}\left(\prod_{i=j+1}^{k} q_j\right)\left(\|\mathbf{X}^j - \mathbf{X}^{j+1}\|^2 + \alpha_j^2\|\mathbf{D}^j - \mathbf{D}^{j+1}\|_M^2\right).$$

Therefore,

$$\min_{j\in[k]} V^j \leq \frac{\prod_{j=0}^{k} q_j}{1 + \sum\limits_{j=1}^{k}\left(\prod_{i=j}^{k} q_j\right)}\frac{V^0}{1-\delta}.$$

Note, that

$$1 \leq q_j \leq \left(\frac{k+1+\beta_1}{k+1}\right)^{2\beta_2}.$$

Define $\overline{\beta_1} := \lceil \beta_1 \rceil$, and recalling

$$\begin{aligned}
\min_{j\in[k]} V^j \leq& \frac{1}{k+1}\left(\frac{\prod_{j=1}^{k}(j+\beta_1+1)}{(k+1)!}\right)^{2\beta_2}\frac{V^0}{1-\delta} \\
\leq& \frac{1}{k+1}\left(\frac{\prod_{j=1}^{k}(j+\overline{\beta_1}+1)}{(k+1)!(\overline{\beta_1}+1)!}\right)^{2\beta_2}\frac{V^0}{1-\delta} \\
=& \frac{1}{k+1}\left(\frac{\prod_{j=k+1}^{k+\overline{\beta_1}+1}(j+\overline{\beta_1})}{(\overline{\beta_1}+1)!}\right)^{2\beta_2}\frac{V^0}{1-\delta} \\
\leq& \frac{1}{k+1}\left(\frac{(k+\overline{\beta_1}+1)^{\overline{\beta_1}}}{(\overline{\beta_1}+1)!}\right)^{2\beta_2}\frac{V^0}{1-\delta} \\
\leq& \frac{c}{(k+1)^{1-2\beta_2\overline{\beta_1}}}V^0.
\end{aligned}$$

The statement of the theorem follows using $\alpha_k \geq \delta/2L$, due to Lemma 3, where L is local smoothness constant.