APPENDIX

Lemma 1. Let $\ker(\mathbf{W}^T) = \ker(\mathbf{Q})$. $\mathbf{QWx} = \mathbf{0}$ if and only if $\mathbf{x} \in \ker(\mathbf{W})$.

I. PROOF OF THEOREM 1

Consider an arbitrary stopping point of the algorithm, that is $\mathbf{x}_{n+1}^{\nu} = \mathbf{x}_n^{\nu} = \mathbf{x}^{\nu}$, $\nabla \mathbf{f}_{n+1} = \nabla \mathbf{f}_n = \nabla \mathbf{f}$, $\mathbf{h}_{n+1}^{\nu} = \mathbf{h}_n^{\nu} = \mathbf{h}^{\nu}$ and $\mathbf{g}_{n+1}^{\nu} = \mathbf{g}_n^{\nu} = \mathbf{g}^{\nu}$. We have

$$\mathbf{Z} = \mathbf{X} - \mathbf{WX} - \mu\left(\nabla \mathbf{f} - \mathbf{G}\right) \tag{1}$$

$$\mathbf{X} = \boldsymbol{\mathcal{P}}_S\left(\mathbf{Z}\right) \tag{2}$$

$$\rho\left[\nabla \mathbf{f} - \mathbf{G} + \frac{1}{\mu}\left(\mathbf{Z} - \mathbf{X}\right)\right] + \alpha\left(\mathbf{H} - \mathbf{G}\right) = \mathbf{O} \tag{3}$$

$$\mathbf{Q}\left(\mathbf{H} - \mathbf{G}\right) = \mathbf{O}. \tag{4}$$

Left multiplying (3) by $\mathbf{Q}$, considering (4), we have

$$\mathbf{Q}(\mathbf{G} - \nabla \mathbf{f}) = \frac{1}{\mu}\mathbf{Q}(\mathbf{Z} - \mathbf{X}). \tag{5}$$

Left multiplying (1) by $\mathbf{Q}$, and applying (5) leads to $\mathbf{QWX} = \mathbf{0}$. Therefore, $\mathbf{X} \in \operatorname{span}\{\mathbf{1}\}$ based on the result of Lemma 1, which means that $\mathbf{x}^{\nu} = \mathbf{x}^*$, $\forall \nu \in \mathcal{V}$. As $\mathbf{X} \in \ker(\mathbf{W})$, (1) reduces to $\mathbf{Z} - \mathbf{X} = \mu(\mathbf{G} - \nabla \mathbf{f})$, which leads to $\mathbf{H} = \mathbf{G}$ by incorporating it into (3). Since (**??**) is designed to preserve the summation of $\mathbf{h}^{\nu}$s, and each element of $\mathbf{H}$ is initialized with zero vector, we have

$$\mathbf{1}^T\mathbf{G} = \mathbf{1}^T\mathbf{H} = \sum_{\nu \in \mathcal{V}} (\mathbf{h}^{\nu})^T = \mathbf{0}^T. \tag{6}$$

From (2), we have $\mathbf{Z} - \mathbf{X} \in \boldsymbol{\partial}\mathbf{I}_S$, consequently, $\mu(\mathbf{G} - \nabla \mathbf{f}) \in \boldsymbol{\partial}\mathbf{I}_S$. As $\boldsymbol{\partial}\mathbf{I}_{S_\nu}$ is a cone, and therefore invariant to scaling, we can write $(\mathbf{G} - \nabla \mathbf{f}) \in \boldsymbol{\partial}\mathbf{I}_S$. Left multiplying by $\mathbf{1}^T$, and moving all the terms to one side, considering (6), we have

$$\mathbf{0} \in \sum_{\nu \in \mathcal{V}} \partial I_{S_\nu}(\mathbf{x}^*) + \nabla f_\nu(\mathbf{x}^*), \tag{7}$$

which shows that $\mathbf{x}^*$ is optimal. ∎

Lemma 2. Consider matrices $\mathbf{R}, \mathbf{P}$ and $\mathbf{S}$, which are respectively defined in (**??**), (**??**) and (**??**). Define a "dual" sequence $\{\boldsymbol{\Lambda}_n\}_{n=-1}^{N-1}$ such that $\boldsymbol{\Lambda}_{N-1} = \boldsymbol{\Lambda}_{-1} = \mathbf{O}$. Suppose that there exists a $C \geq 0$ such that for every $\beta > 0$, the following system of equations

$$\boldsymbol{\Lambda}_{n-1} - \mathbf{R}^T\boldsymbol{\Lambda}_n + (\mathbf{S} + C\delta_{n,0}\mathbf{I})\boldsymbol{\Psi}_n = \mathbf{O} \qquad n = 0, 1, \ldots, N-1 \tag{8}$$

$$\mathbf{P}^T\boldsymbol{\Lambda}_n = \beta\mathbf{X}_{n+2} \qquad n = 0, 1, \ldots, N-2 \tag{9}$$

has no non-zero solution for $\{\boldsymbol{\Psi}_n, \boldsymbol{\Lambda}_n, \mathbf{X}_{n+2}\}$. Then, $\mathbf{A}_n \geq -C\|\boldsymbol{\Psi}_0\|_{\mathrm{F}}^2$ always holds true.

Proof. Note that the claim is equivalent to the statement that zero is the optimal value for the optimization problem

$$\min_{\{\boldsymbol{\Psi}_n\}_{n=0}^{N-1}, \{\mathbf{X}_{n+2}\}_{n=0}} \sum_{n=0}^{N-1} \langle \boldsymbol{\Psi}_n, \mathbf{S}\boldsymbol{\Psi}_n \rangle + C\|\boldsymbol{\Psi}_0\|^2$$
$$\text{s.t}$$
$$\boldsymbol{\Psi}_{n+1} = \mathbf{R}\boldsymbol{\Psi}_n + \mathbf{P}\mathbf{X}_{n+2}, \quad n = 0, 1, \ldots, N-2 \tag{10}$$

If the claim does not hold, the optimization is unbounded and the following restricted optimization will achieve a strictly negative optimal value at a non-zero solution:

$$\min_{\{\boldsymbol{\Psi}_n\}_{n=0}^{N-1}, \{\mathbf{X}_{n+2}\}_{n=0}} \sum_{n=0}^{N-1} \langle \boldsymbol{\Psi}_n, \mathbf{S}\boldsymbol{\Psi}_n \rangle + C\|\boldsymbol{\Psi}_0\|^2$$
$$\text{s.t}$$
$$\boldsymbol{\Psi}_{n+1} = \mathbf{R}\boldsymbol{\Psi}_n + \mathbf{P}\mathbf{X}_{n+2}, \quad n = 0, 1, \ldots, N-2$$
$$\sum_{n=0}^{N-2} \|\mathbf{X}_{n+2}\|_{\mathrm{F}}^2 \leq 1 \tag{11}$$

Such a solution satisfies the KKT condition, which coincides with (8), (9) where $\{\boldsymbol{\Lambda}_n\}, \beta \geq 0$ are dual (Lagrangian) multipliers corresponding to the constraints. We also observe that the optimal value at this point is given by $-\beta \sum_{n=0}^{N-2} \|\mathbf{X}_{n+2}\|_{\mathrm{F}}^2$. This shows that $\beta > 0$. This contradicts the assumption that such a point does not exists and completes the proof. ∎

We may further simplify the conditions in Lemma 2 by the following result:

Lemma 3. Consider $\mathbf{F}(z, \beta)$ defined in (**??**) with a complex value z and a real value β. The condition of Lemma 2 is satisfied if the following limit exists.

$$\lim_{z \to 0} \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{F}^{-1}(z, \beta) \begin{bmatrix} -C\mathbf{I} \\ \mathbf{I} \\ \mathbf{0} \end{bmatrix} \tag{12}$$

Proof. With an abuse of notation, define the z-transforms

$$\boldsymbol{\Psi}(z) = \sum_{n=0}^{N-1} \boldsymbol{\Psi}_n z^{-n}, \quad \boldsymbol{\Lambda}(z) = \sum_{n=0}^{N-2} \boldsymbol{\Lambda}_n z^{-n}$$

$$\mathbf{U}(z) = \sum_{n=0}^{N-2} \mathbf{X}_{n+2} z^{-n}$$

Then, we have

$$(z^{-1}\mathbf{I} - \mathbf{R}^T)\boldsymbol{\Lambda}(z) + \mathbf{S}\boldsymbol{\Psi}(z) + C\boldsymbol{\Psi}_0 = 0 \tag{13}$$

$$(z\mathbf{I} - \mathbf{R})\boldsymbol{\Psi}(z) - \boldsymbol{\Psi}_0 + \mathbf{R}\boldsymbol{\Psi}_{N-1}z^{n-1} + \mathbf{P}\mathbf{U}(z) = 0 \tag{14}$$

$$\mathbf{P}^T\boldsymbol{\Lambda}(z) = -\beta\mathbf{U}(z) \tag{15}$$

which can also be written as

$$\mathbf{F}(z, \beta) \begin{bmatrix} \boldsymbol{\Psi}(z) \\ \boldsymbol{\Lambda}(z) \\ \mathbf{U}(z) \end{bmatrix} = \begin{bmatrix} -C\boldsymbol{\Psi}_0 \\ \boldsymbol{\Psi}_0 - \mathbf{R}\boldsymbol{\Psi}_{N-1}z^{N-1} \\ \mathbf{0} \end{bmatrix} \tag{16}$$

Note that $\mathbf{F}(z, \beta)$ may be rank-deficient at a finite number of points. Hence, there exists a sufficiently small simple loop $\mathcal{C}$ around $z = 0$ such that $\mathbf{F}$ is invertible on an inside it except at $z = 0$. In this region we have

$$\boldsymbol{\Psi}(z) = \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{F}^{-1}\mathbf{A}, \tag{17}$$

where $\mathbf{A}$ is

$$\begin{bmatrix} -C\mathbf{I} \\ \mathbf{I} \\ \mathbf{0} \end{bmatrix} \boldsymbol{\Psi}_0 + \begin{bmatrix} \mathbf{0} \\ -\mathbf{R} \\ \mathbf{0} \end{bmatrix} z^{N-1}\boldsymbol{\Psi}_{N-1}$$

On the other hand,

$$2\pi j \mathbf{\Psi}_0 = \oint_{\mathcal{C}} \frac{1}{z}\mathbf{\Psi}(z)\mathrm{d}z \tag{18}$$

Further from the Cauchy integral formula for sufficiently large N, we have

$$\oint_{\mathcal{C}} \frac{1}{z} z^{N-1}\mathbf{F}^{-1}(z,\beta)\mathrm{d}z = 2\pi j \lim_{z\to 0} z^{N-1}\mathbf{F}^{-1}(z,\beta) = \mathbf{0} \tag{19}$$

We conclude that

$$2\pi j \Psi_0 = \lim_{z\to 0} \left[\begin{array}{ccc} \mathbf{I} & \mathbf{0} & \mathbf{0} \end{array}\right] \mathbf{F}^{-1} \left[\begin{array}{c} -C\mathbf{I} \\ \mathbf{I} \\ \mathbf{0} \end{array}\right] \mathbf{\Psi}_0 \tag{20}$$

Note that by the assumption the right hand side is finite and real. However, the left hand side is imaginary, showing that $\mathbf{\Psi}_0 = 0$, which completes the proof. ■

II. PROOF OF THEOREM 2

We start by defining

$$F^v(\mathbf{x}) = f_v(\mathbf{x}) - f_v(\mathbf{x}^*) - \langle \nabla f_v(\mathbf{x}^*), \mathbf{x} - \mathbf{x}^* \rangle \tag{21}$$

and

$$F_n^v = F^v(\mathbf{x}_n^v) \tag{22}$$

Note that from the convexity of f_v, the values of $F^v(\mathbf{x})$, particularly F_n^v are non-negative. From convexity, we also conclude that

$$\begin{aligned} &F_n^v + \langle \nabla f_v(\mathbf{x}^*) - \nabla f_v(\mathbf{x}_n^v),\ \mathbf{x}_n^v - \mathbf{x}^* \rangle \\ &= f_v(\mathbf{x}_n^v) - f_v(\mathbf{x}^*) + \langle \nabla f_v(\mathbf{x}_n^v),\ \mathbf{x}^* - \mathbf{x}_n^v \rangle \le 0 \end{aligned} \tag{23}$$

From the $L-$smoothness property of f_v, we also obtain

$$\begin{aligned} &F_{n+1}^v - F_n^v - \langle \nabla f_v(\mathbf{x}_n^v) - \nabla f_v(\mathbf{x}^*), \mathbf{x}_{n+1}^v - \mathbf{x}_n^v \rangle \\ &= f_v(\mathbf{x}_{n+1}^v) - f_v(\mathbf{x}_n^v) - \langle \nabla f_v(\mathbf{x}_n^v), \mathbf{x}_{n+1}^v - \mathbf{x}_n^v \rangle \\ &\le \frac{L}{2}\left\| \mathbf{x}_{n+1}^v - \mathbf{x}_n^v \right\|^2 \end{aligned} \tag{24}$$

Adding (23) to (24) yields:

$$\begin{aligned} F_{n+1}^v + \left\langle \nabla f_v(\mathbf{x}^*) - \nabla f_v(\mathbf{x}_n^v),\ \mathbf{x}_{n+1}^v - \mathbf{x}^* \right\rangle& \\ - \frac{L}{2}\left\| \mathbf{x}_{n+1}^v - \mathbf{x}_n^v \right\|^2 \le 0& \end{aligned} \tag{25}$$

Now, we define

$$T^v(\mathbf{x}) = -\langle \mathbf{n}^v, \mathbf{x} - \mathbf{x}^* \rangle \tag{26}$$

and $T_n^v = T^v(\mathbf{x}_n^v)$. The fact that $\mathbf{x}_{n+1}^v \in S_v$ yields $T_{n+1}^v \ge 0$. Note that from $\mathbf{x}_{n+1}^v = \mathcal{P}_{S_v}(\mathbf{z}_n^v)$ and the fact that $\mathbf{x}^* \in S_v$, we have

$$\langle \mathbf{x}^* - \mathbf{x}_{n+1}^v, \mathbf{z}_n^v - \mathbf{x}_{n+1}^v \rangle \le 0 \tag{27}$$

which can also be written as

$$\mu T_{n+1}^v + \langle \mathbf{x}^* - \mathbf{x}_{n+1}^v, \mathbf{z}_n^v - \mathbf{x}_{n+1}^v - \mu \mathbf{n}^v \rangle \le 0 \tag{28}$$

Multiplying (25) to μ, adding to (28), plugging the definition of $\mathbf{z}_n^v$ in and summing over $v \in G$ and $n = 0, 1, \ldots, N-1$ we obtain

$$\begin{aligned} &\mu \sum_{n\in[N],v} F_{n+1}^v + T_{n+1}^v - \frac{L\mu}{2} \sum_{n\in[N],v} \left\| \mathbf{x}_{n+1}^v - \mathbf{x}_n^v \right\|^2 \\ &+ \sum_{n\in[N],v} \Big\langle \mathbf{x}^* - \mathbf{x}_{n+1}^v, \mathbf{x}_n^v - \sum_u w_{vu}\mathbf{x}_n^u - \mathbf{x}_{n+1}^v \\ &\qquad + \mu(\mathbf{g}_n^v - \nabla f_v(\mathbf{x}^*) - \mathbf{n}^v) \Big\rangle \le 0 \end{aligned} \tag{29}$$

We also replace the expression of $\mathbf{z}_n^v$ in the dynamics of $\mathbf{g}_n^v$ leading to

$$\mathbf{g}_{n+1}^v = \mathbf{g}_n^v + \frac{\rho}{\mu}\left(\mathbf{x}_n^v - \sum_u w_{vu}\mathbf{x}_n^u - \mathbf{x}_{n+1}^v \right) + \alpha \boldsymbol{\delta}_n^v, \tag{30}$$

where $\boldsymbol{\delta}_n^v = \mathbf{h}_n^v - \mathbf{g}_n^v$ that follows the following dynamics:

$$\begin{aligned} \boldsymbol{\delta}_{n+1}^v =\ &(1-\alpha)\boldsymbol{\delta}_n^v - \sum_u q_{vu}\boldsymbol{\delta}_n^u \\ &- \frac{\rho}{\mu}\left(\mathbf{x}_n^v - \sum_u w_{vu}\mathbf{x}_n^u - \mathbf{x}_{n+1}^v \right) \end{aligned} \tag{31}$$

For simplicity, we define $\tilde{\mathbf{x}}_n^v = \mathbf{x}_n^v - \mathbf{x}^*$ and $\tilde{\mathbf{g}}_n^v = \mathbf{g}_n^v - \nabla f_v(\mathbf{x}^*) - \mathbf{n}^v$ and rewrite (29),(30) and (31) as

$$\begin{aligned} &\sum_{n=0}^{N-1} \left(\mu \sum_v F_{n+1}^v + T_{n+1}^v + \frac{\eta}{2} \sum_{u,v} \|\mathbf{x}_{n+1}^u - \mathbf{x}_{n+1}^v\|^2 \right) \\ &- \sum_{n=0}^{N-1} \sum_v \left\langle \tilde{\mathbf{x}}_{n+1}^v, \tilde{\mathbf{x}}_n^v - \sum_u w_{vu}\tilde{\mathbf{x}}_n^u - \tilde{\mathbf{x}}_{n+1}^v + \mu \tilde{\mathbf{g}}_n^v \right\rangle \\ &- \frac{L\mu}{2} \sum_{n=0}^{N-1} \sum_v \left\| \tilde{\mathbf{x}}_{n+1}^v - \tilde{\mathbf{x}}_n^v \right\|^2 \\ &- \frac{\eta}{2} \sum_{n=0}^{N-1} \sum_{u,v} \|\tilde{\mathbf{x}}_{n+1}^u - \tilde{\mathbf{x}}_{n+1}^v\|^2 \le 0 \end{aligned} \tag{32}$$

$$\tilde{\mathbf{g}}_{n+1}^v = \tilde{\mathbf{g}}_n^v + \frac{\rho}{\mu}\left(\tilde{\mathbf{x}}_n^v - \sum_u w_{vu}\tilde{\mathbf{x}}_n^u - \tilde{\mathbf{x}}_{n+1}^v \right) + \alpha \boldsymbol{\delta}_n^v, \tag{33}$$

$$\begin{aligned} \boldsymbol{\delta}_{n+1}^v = (1-\alpha)\boldsymbol{\delta}_n^v - \sum_u q_{vu}\boldsymbol{\delta}_n^u& \\ - \frac{\rho}{\mu}\left(\tilde{\mathbf{x}}_n^v - \sum_u w_{vu}\tilde{\mathbf{x}}_n^u - \tilde{\mathbf{x}}_{n+1}^v \right)& \end{aligned} \tag{34}$$

where in the first inequality we also add and remove the term $\frac{\eta}{2}\sum_{u,v} \|\mathbf{x}_{n+1}^u - \mathbf{x}_{n+1}^v\|^2$. In the following, we will consider the last three summations in (32) as A_N. We show an asymptotic lower bound for A_N, i.e show that there exists a constant C only depending on the initial values such that for sufficiently large N, $A_N \ge -C$. Note that since $A_N \le 0$, we must have $C \ge 0$. Then, we conclude from (32) that

$$\sum_{n=0}^{N-1} \left(\mu \sum_v F_{n+1}^v + T_{n+1}^v + \frac{\eta}{2} \sum_{u,v} \|\mathbf{x}_{n+1}^u - \mathbf{x}_{n+1}^v\|^2 \right) \le C \tag{35}$$

Defining $\bar{\mathbf{x}}_N^v = \frac{1}{N}\sum_{n=0}^{N-1} \mathbf{x}_N^v$ and noting that each term in the summation over n is a fixed convex function of $\{\mathbf{x}_{n+1}^v\}_v$, we may recall Jensen's inequality to conclude

$$\mu\sum_v F^v(\bar{\mathbf{x}}_N^v) + T^v(\bar{\mathbf{x}}_N^v) + \frac{\eta}{2}\sum_{u,v}\|\bar{\mathbf{x}}_N^u - \bar{\mathbf{x}}_N^v\|^2 \leq \frac{C}{N} \quad (36)$$

Defining $\bar{\mathbf{x}}_N = \frac{1}{M}\sum_v \bar{\mathbf{x}}_N^v$, we conclude that $\|\bar{\mathbf{x}}_N - \bar{\mathbf{x}}_N^u\|^2 = O(\frac{1}{N})$. Since $\bar{\mathbf{x}}_N^u \in S_u$, we also conclude that $\text{dist}^2(\bar{\mathbf{x}}_N, S_u) = O(\frac{1}{N})$. Finally,

$$\begin{aligned}
&\left|\sum_v f_v(\bar{\mathbf{x}}_N^v) - \sum_v f_v(\mathbf{x}^*)\right| \\
&\leq \frac{C}{\mu N} + \sum_v |\langle \mathbf{n}^v + \nabla f_v(\mathbf{x}^*), \bar{\mathbf{x}}_N^v - \bar{\mathbf{x}}_N\rangle| \\
&\leq \frac{C}{\mu N} + \sqrt{\sum_v \|\mathbf{n}^v + \nabla f_v(\mathbf{x}^*)\|^2}\sqrt{\sum_v \|\bar{\mathbf{x}}_N^v - \bar{\mathbf{x}}_N\|^2} \\
&= O(\frac{1}{\sqrt{N}})
\end{aligned}$$

which completes the proof.

Bound on A_N: To find the bound C, we start by simplifying the notation in (32), (33) and (34). Let us introduce

$$\mathbf{\Psi}_n = \begin{bmatrix} \tilde{\mathbf{X}}_{n+1} & \tilde{\mathbf{X}}_n & \tilde{\mathbf{G}}_n & \mathbf{\Delta}_n \end{bmatrix}^T, \quad (37)$$

where $\tilde{\mathbf{X}}_n, \tilde{\mathbf{G}}_n, \mathbf{\Delta}_n$ are matrices with $\tilde{\mathbf{x}}_n^v, \tilde{\mathbf{g}}_n^v, \boldsymbol{\delta}_n^v$ as their v^{th} row, respectively. We may write (33) and (34) as

$$\mathbf{\Psi}_{n+1} = \mathbf{R}\mathbf{\Psi}_n + \mathbf{P}\mathbf{X}_{n+2} \qquad n = 0, \dots, N-2, \quad (38)$$

where $\mathbf{P}$ and $\mathbf{R}$ are defined as

$$\mathbf{R} = \begin{bmatrix} \mathbf{O} & \mathbf{O} & \mathbf{O} & \mathbf{O} \\ \mathbf{I} & \mathbf{O} & \mathbf{O} & \mathbf{O} \\ -\frac{\rho}{\mu}\mathbf{I} & \frac{\rho}{\mu}(\mathbf{I}-\mathbf{W}) & \mathbf{I} & \alpha\mathbf{I} \\ \frac{\rho}{\mu}\mathbf{I} & -\frac{\rho}{\mu}(\mathbf{I}-\mathbf{W}) & \mathbf{O} & (1-\alpha)\mathbf{I}-\mathbf{Q} \end{bmatrix} \quad (39)$$

$$\mathbf{P} = \begin{bmatrix} \mathbf{I} & \mathbf{O} & \mathbf{O} & \mathbf{O} \end{bmatrix}^T. \quad (40)$$

We also have

$$A_N = \sum_{n=0}^{N-1} \langle \mathbf{\Psi}_n, \mathbf{S}\mathbf{\Psi}_n\rangle \quad (41)$$

where $\mathbf{S} = [\mathbf{S}_1 \ \mathbf{S}_2]$ is computed as

$$\mathbf{S}_1 = \begin{bmatrix} \left(1-\frac{L\mu}{2}\right)\mathbf{I} - M\eta\left(\mathbf{I} - \frac{1}{M}\mathbf{1}\mathbf{1}^T\right) \\ -\frac{1}{2}(\mathbf{I}-\mathbf{W}^T) + \frac{L\mu}{2}\mathbf{I} \\ -\frac{\mu}{2}\mathbf{I} \\ \mathbf{0} \end{bmatrix}$$

$$\mathbf{S}_2 = \begin{bmatrix} -\frac{1}{2}(\mathbf{I}-\mathbf{W}) + \frac{L\mu}{2}\mathbf{I} & -\frac{\mu}{2}\mathbf{I} & \mathbf{0} \\ -\frac{L\mu}{2}\mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

The proof is completed using the outcome of Lemma 2. ■