9. Proof of Lemma 1

Let $(\boldsymbol{x}^*, \boldsymbol{\alpha}^*)$ be a solution to the VI in (17). We want to show that $(\boldsymbol{x}^*, \boldsymbol{\alpha}^*) \in \mathcal{N}_{\text{opt}}$. First, since the VI holds for all $\boldsymbol{x}, \boldsymbol{\alpha}$ we can pick $\boldsymbol{\alpha} = \boldsymbol{\alpha}^*$, so for all $\boldsymbol{x} \in \mathcal{X}$

$$\langle \boldsymbol{x} - \boldsymbol{x}^*, F(\boldsymbol{x}^*, \boldsymbol{\alpha}^*) \rangle \leq 0. \tag{21}$$

Then by Proposition 1.4.2 in (Facchinei & Pang, 2007), $\boldsymbol{x}^*$ is a NE (note that because $F(\boldsymbol{x}, \boldsymbol{\alpha}^*)$ is strongly monotone in $\boldsymbol{x}$, then $r_n(\boldsymbol{x}_n, \boldsymbol{x}_{-n})$ is concave in $\boldsymbol{x}_n$ for each n and $\boldsymbol{x}_{-n}$).

For any k, we can pick $\boldsymbol{x} = \boldsymbol{x}^*$ and $\boldsymbol{\alpha} = \boldsymbol{\alpha}_0$ such that $\alpha_0^l = \alpha^{*l}$ for all $l \neq k$ and $\alpha_0^k = \alpha^{*k} + \varepsilon$ for some $\varepsilon > 0$, and get from the VI in (17) that

$$\varepsilon \sum_{n=1}^{N} \left(x_n^{*k} - l_k^* \right) \leq 0 \Longrightarrow \sum_{n=1}^{N} x_n^{*k} \leq l_k^*. \tag{22}$$

Now let k be a coordinate for which $\alpha^{k*} > 0$ (if it exists) and pick $\boldsymbol{x} = \boldsymbol{x}^*$ and $\boldsymbol{\alpha} = \boldsymbol{\alpha}_0$ such that $\alpha_0^l = \alpha^{*l}$ for all $l \neq k$ and $\alpha_0^k = 0$. Then the VI in (17) gives

$$\alpha^{*k} \sum_{n=1}^{N} \left(x_n^{*k} - l_k^* \right) \geq 0 \tag{23}$$

so from (22) and (23) we conclude that $\sum_{n=1}^{N} x_n^{*k} = l_k^*$. Hence for every $\boldsymbol{\alpha}^* \in \mathcal{A}^*$ we have that for all k

$$\sum_{n=1}^{N} x_n^{*k} = l_k^* \text{ or } \left[\sum_{n=1}^{N} x_n^{*k} < l_k^* \text{ and } \alpha_k = 0 \right] \tag{24}$$

so $(\boldsymbol{x}^*, \boldsymbol{\alpha}^*) \in \mathcal{N}_{\text{opt}}$.

Now let $(\boldsymbol{x}^*(\boldsymbol{\alpha}^*), \boldsymbol{\alpha}^*) \in \mathcal{N}_{\text{opt}}$. We want to show that $(\boldsymbol{x}^*(\boldsymbol{\alpha}^*), \boldsymbol{\alpha}^*)$ solves the VI in (17). Since $\boldsymbol{\alpha}^*$ satisfies (24) for all k then for every $\boldsymbol{\alpha} \in \mathbb{R}_+^K$

$$\left\langle \boldsymbol{\alpha} - \boldsymbol{\alpha}^*, \sum_{n=1}^{N} \boldsymbol{x}_n^*(\boldsymbol{\alpha}^*) - \boldsymbol{l}^* \right\rangle = \left\langle \boldsymbol{\alpha}, \sum_{n=1}^{N} \boldsymbol{x}_n^*(\boldsymbol{\alpha}^*) - \boldsymbol{l}^* \right\rangle \leq 0. \tag{25}$$

Additionally, Since $\boldsymbol{x}^*(\boldsymbol{\alpha}^*)$ is a NE then by Proposition 1.4.2 in (Facchinei & Pang, 2007) we have for all $\boldsymbol{x} \in \mathcal{X}$

$$\langle \boldsymbol{x} - \boldsymbol{x}^*, F(\boldsymbol{x}^*(\boldsymbol{\alpha}^*), \boldsymbol{\alpha}^*) \rangle \leq 0. \tag{26}$$

Hence $(\boldsymbol{x}^*(\boldsymbol{\alpha}^*), \boldsymbol{\alpha}^*)$ is a solution to the VI.

10. Proof of Lemma 2

We start by showing that a large enough $\boldsymbol{\alpha}_0$ leads to a NE where the total loads are below $\boldsymbol{l}^*$. Let

$$\mathcal{U}_n = \left\{ \boldsymbol{x}_n \,\middle|\, 0 \leq x_n^k < \frac{l_k^*}{N}, \forall k \right\} \tag{27}$$

and let $\mathcal{X}' = \mathcal{X} \setminus \mathcal{U}_1 \times \ldots \times \mathcal{U}_N$, which is a closed set as the difference of a closed and an open set. Since $r_n(\boldsymbol{x})$ is continuous on the compact set $\mathcal{X}'$ then $\max_{n, \boldsymbol{x} \in \mathcal{X}'} r_n(\boldsymbol{x}) \leq M$ for some $M > 0$. If we choose $\alpha_0^k \geq 2N \frac{M}{l_k^*}$ for all k, then for some player n and for all $\boldsymbol{x} \in \mathcal{X}'$

$$u_n(\boldsymbol{x}) = r_n(\boldsymbol{x}) - \sum_{k=1}^{K} \alpha_0^k x_n^k \leq M \left(1 - 2 \sum_{k=1}^{K} \frac{N}{l^{k*}} x_n^k \right) < 0. \tag{28}$$

Hence, no $\boldsymbol{x} \in \mathcal{X}'$ is a NE since by switching to $\boldsymbol{x}_n = \boldsymbol{0}$ player n receives $u_n(\boldsymbol{x}) = 0$. We conclude that $\boldsymbol{x}^*(\boldsymbol{\alpha}_0) \in \mathcal{U}_1 \times \ldots \times \mathcal{U}_N$, so $\sum_{n=1}^{N} x_n^{*k}(\boldsymbol{\alpha}_0) \leq l_k^*$ for all k.

Next we use this $\boldsymbol{\alpha}_0$ to argue about the set of solutions to our VI in (17). For each $\boldsymbol{\alpha}$ we have

$$\begin{aligned} &\langle \boldsymbol{x} - \boldsymbol{x}^*(\boldsymbol{\alpha}_0), F(\boldsymbol{x}, \boldsymbol{\alpha}) \rangle + \left\langle \boldsymbol{\alpha} - \boldsymbol{\alpha}_0, \sum_{n=1}^{N} \boldsymbol{x}_n - \boldsymbol{l}^* \right\rangle \underset{(a)}{\leq} \\ &\langle \boldsymbol{x} - \boldsymbol{x}^*(\boldsymbol{\alpha}_0), F(\boldsymbol{x}^*(\boldsymbol{\alpha}_0), \boldsymbol{\alpha}) \rangle + \left\langle \boldsymbol{\alpha} - \boldsymbol{\alpha}_0, \sum_{n=1}^{N} \boldsymbol{x}_n - \boldsymbol{l}^* \right\rangle = \\ &\langle \boldsymbol{x} - \boldsymbol{x}^*(\boldsymbol{\alpha}_0), F(\boldsymbol{x}^*(\boldsymbol{\alpha}_0), \boldsymbol{\alpha}_0) \rangle + \left\langle \sum_{n=1}^{N} (\boldsymbol{x}_n - \boldsymbol{x}_n^*(\boldsymbol{\alpha}_0)), \boldsymbol{\alpha}_0 - \boldsymbol{\alpha} \right\rangle + \\ &\left\langle \boldsymbol{\alpha} - \boldsymbol{\alpha}_0, \sum_{n=1}^{N} \boldsymbol{x}_n - \boldsymbol{l}^* \right\rangle \underset{(b)}{\leq} \left\langle \boldsymbol{\alpha} - \boldsymbol{\alpha}_0, \sum_{n=1}^{N} \boldsymbol{x}_n^*(\boldsymbol{\alpha}_0) - \boldsymbol{l}^* \right\rangle \end{aligned} \tag{29}$$

where (a) uses the monotonicity of $F(\boldsymbol{x}, \boldsymbol{\alpha})$ in $\boldsymbol{x}$ and (b) follows since $\boldsymbol{x}^*(\boldsymbol{\alpha}_0)$ is a NE (Proposition 1.4.2 in (Facchinei & Pang, 2007)). Hence the set

$$L_{\geq} = \left\{ (\boldsymbol{x}, \boldsymbol{\alpha}) \in \mathcal{X} \times \mathbb{R}_+^K \,\middle|\, \langle \boldsymbol{x} - \boldsymbol{x}^*(\boldsymbol{\alpha}_0), F(\boldsymbol{x}, \boldsymbol{\alpha}) \rangle + \left\langle \boldsymbol{\alpha} - \boldsymbol{\alpha}_0, \sum_{n=1}^{N} \boldsymbol{x}_n - \boldsymbol{l}^* \right\rangle \geq 0 \right\} \tag{30}$$

is bounded, since (29) shows that

$$L_{\geq} \subseteq \{ (\boldsymbol{x}, \boldsymbol{\alpha}) \,|\, \boldsymbol{x} \in \mathcal{X}, \boldsymbol{\alpha} \in \mathcal{C} \} \tag{31}$$

where $\mathcal{C}$ is the following bounded convex polytope

$$\mathcal{C} = \left\{ \boldsymbol{\alpha} \in \mathbb{R}_+^K \,\middle|\, \left\langle \boldsymbol{\alpha} - \boldsymbol{\alpha}_0, \sum_{n=1}^{N} \boldsymbol{x}_n^*(\boldsymbol{\alpha}_0) - \boldsymbol{l}^* \right\rangle \geq 0 \right\} = \left\{ \boldsymbol{\alpha} \in \mathbb{R}_+^K \,\middle|\, \langle \boldsymbol{\alpha}, \boldsymbol{v} \rangle \leq \underbrace{\langle \boldsymbol{\alpha}_0, \boldsymbol{v} \rangle}_{\geq 0} \right\} \tag{32}$$

where $\boldsymbol{v}=\sum_{n=1}^{N}\left(\boldsymbol{l}^{*}-\boldsymbol{x}_{n}^{*}\left(\boldsymbol{\alpha}_{0}\right)\right)\geq 0$. Therefore according to Proposition 2.2.3 in (Facchinei & Pang, 2007) the set of solutions to the VI in (17) is non-empty and compact, which by Lemma 1 is $\mathcal{N}_{\text{opt}}=\mathcal{X}^{*}\times\mathcal{A}^{*}$.

11. Proof of Lemma 3

Note that $\mathcal{X}\triangleq\mathcal{X}_{1}\times\ldots\times\mathcal{X}_{N}$ is closed and convex since $\mathcal{X}_{n}$ is closed convex for each n. Also note that $F\left(\boldsymbol{x},\boldsymbol{\alpha}\right)$ is Lipschitz continuous in $\boldsymbol{x}$ since it is continuously differentiable on the closed $\mathcal{X}$. Then since $F\left(\boldsymbol{x},\boldsymbol{\alpha}\right)$ is strongly monotone on $\mathcal{X}$ (given $\boldsymbol{\alpha}$), Theorem 2.3.3 in (Facchinei & Pang, 2007) states that for all $\boldsymbol{x}\in\mathcal{X}$, for some $L_{0}>0$

$$\left\|\boldsymbol{x}-\boldsymbol{x}^{*}\left(\boldsymbol{\alpha}\right)\right\|\leq L_{0}\left\|\boldsymbol{x}-\Pi_{\mathcal{X}}\left(\boldsymbol{x}-F\left(\boldsymbol{x},\boldsymbol{\alpha}\right)\right)\right\|. \tag{33}$$

Hence for $\boldsymbol{x}^{*}\left(\boldsymbol{\alpha}_{2}\right),\boldsymbol{x}^{*}\left(\boldsymbol{\alpha}_{1}\right)$ we have

$$\begin{aligned}
&\left\|\boldsymbol{x}^{*}\left(\boldsymbol{\alpha}_{2}\right)-\boldsymbol{x}^{*}\left(\boldsymbol{\alpha}_{1}\right)\right\|\leq\\
&L_{0}\left\|\boldsymbol{x}^{*}\left(\boldsymbol{\alpha}_{2}\right)-\Pi_{\mathcal{X}}\left(\boldsymbol{x}^{*}\left(\boldsymbol{\alpha}_{2}\right)-F\left(\boldsymbol{x}^{*}\left(\boldsymbol{\alpha}_{2}\right),\boldsymbol{\alpha}_{1}\right)\right)\right\|=\\
&L_{0}\Bigg\|\boldsymbol{x}^{*}\left(\boldsymbol{\alpha}_{2}\right)-\Pi_{\mathcal{X}}\left(\boldsymbol{x}^{*}\left(\boldsymbol{\alpha}_{2}\right)-F\left(\boldsymbol{x}^{*}\left(\boldsymbol{\alpha}_{2}\right),\boldsymbol{\alpha}_{1}\right)\right)+\\
&\Pi_{\mathcal{X}}\left(\boldsymbol{x}^{*}\left(\boldsymbol{\alpha}_{2}\right)-F\left(\boldsymbol{x}^{*}\left(\boldsymbol{\alpha}_{2}\right),\boldsymbol{\alpha}_{2}\right)\right)-\\
&\Pi_{\mathcal{X}}\left(\boldsymbol{x}^{*}\left(\boldsymbol{\alpha}_{2}\right)-F\left(\boldsymbol{x}^{*}\left(\boldsymbol{\alpha}_{2}\right),\boldsymbol{\alpha}_{2}\right)\right)\Bigg\|\underset{(a)}{=}\\
&L_{0}\Bigg\|\Pi_{\mathcal{X}}\left(\boldsymbol{x}^{*}\left(\boldsymbol{\alpha}_{2}\right)-F\left(\boldsymbol{x}^{*}\left(\boldsymbol{\alpha}_{2}\right),\boldsymbol{\alpha}_{2}\right)\right)-\\
&\Pi_{\mathcal{X}}\left(\boldsymbol{x}^{*}\left(\boldsymbol{\alpha}_{2}\right)-F\left(\boldsymbol{x}^{*}\left(\boldsymbol{\alpha}_{2}\right),\boldsymbol{\alpha}_{1}\right)\right)\Bigg\|\leq\\
&L_{0}\left\|F\left(\boldsymbol{x}^{*}\left(\boldsymbol{\alpha}_{2}\right),\boldsymbol{\alpha}_{1}\right)-F\left(\boldsymbol{x}^{*}\left(\boldsymbol{\alpha}_{2}\right),\boldsymbol{\alpha}_{2}\right)\right\|=\\
&\sqrt{N}L_{0}\left\|\boldsymbol{\alpha}_{2}-\boldsymbol{\alpha}_{1}\right\|
\end{aligned} \tag{34}$$

where in (a) we used that

$$\boldsymbol{x}^{*}\left(\boldsymbol{\alpha}_{2}\right)-\Pi_{\mathcal{X}}\left(\boldsymbol{x}^{*}\left(\boldsymbol{\alpha}_{2}\right)-F\left(\boldsymbol{x}^{*}\left(\boldsymbol{\alpha}_{2}\right),\boldsymbol{\alpha}_{2}\right)\right)=\mathbf{0} \tag{35}$$

which follows from Proposition 1.5.8 in (Facchinei & Pang, 2007).

12. Proof of Lemma 4

Let $\boldsymbol{\alpha}_{1},\boldsymbol{\alpha}_{2}\in\mathbb{R}_{+}^{K}$. Let $\boldsymbol{x}_{1}^{*}=\boldsymbol{x}^{*}\left(\boldsymbol{\alpha}_{1}\right)$ and $\boldsymbol{x}_{2}^{*}=\boldsymbol{x}^{*}\left(\boldsymbol{\alpha}_{2}\right)$. Since $\boldsymbol{x}_{1}^{*}$ is a NE, we have for every $\boldsymbol{x}\in\mathcal{X}$ that (see Proposition 1.4.2 in (Facchinei & Pang, 2007)):

$$\left\langle\boldsymbol{x}-\boldsymbol{x}_{1}^{*},F\left(\boldsymbol{x}_{1}^{*},\boldsymbol{\alpha}_{1}\right)\right\rangle\leq 0 \tag{36}$$

so for $\boldsymbol{x}=\boldsymbol{x}_{2}^{*}$

$$\left\langle\boldsymbol{x}_{2}^{*}-\boldsymbol{x}_{1}^{*},F\left(\boldsymbol{x}_{1}^{*},\boldsymbol{\alpha}_{1}\right)\right\rangle\leq 0. \tag{37}$$

Since $\boldsymbol{x}_{2}^{*}$ is a NE, we have for $\boldsymbol{x}\in\mathcal{X}$ that

$$\left\langle\boldsymbol{x}-\boldsymbol{x}_{2}^{*},F\left(\boldsymbol{x}_{2}^{*},\boldsymbol{\alpha}_{2}\right)\right\rangle\leq 0 \tag{38}$$

so for $\boldsymbol{x}=\boldsymbol{x}_{1}^{*}$

$$\left\langle\boldsymbol{x}_{1}^{*}-\boldsymbol{x}_{2}^{*},F\left(\boldsymbol{x}_{2}^{*},\boldsymbol{\alpha}_{2}\right)\right\rangle\leq 0. \tag{39}$$

By adding (37) and (39) we obtain

$$\left\langle\boldsymbol{x}_{2}^{*}-\boldsymbol{x}_{1}^{*},F\left(\boldsymbol{x}_{2}^{*},\boldsymbol{\alpha}_{2}\right)-F\left(\boldsymbol{x}_{1}^{*},\boldsymbol{\alpha}_{1}\right)\right\rangle\geq 0. \tag{40}$$

Then

$$\begin{aligned}
&-\lambda\left\|\boldsymbol{x}^{*}\left(\boldsymbol{\alpha}_{2}\right)-\boldsymbol{x}^{*}\left(\boldsymbol{\alpha}_{1}\right)\right\|^{2}=-\lambda\left\|\boldsymbol{x}_{2}^{*}-\boldsymbol{x}_{1}^{*}\right\|^{2}\underset{(a)}{\geq}\\
&\left\langle\boldsymbol{x}_{2}^{*}-\boldsymbol{x}_{1}^{*},F\left(\boldsymbol{x}_{2}^{*},\boldsymbol{\alpha}_{1}\right)-F\left(\boldsymbol{x}_{1}^{*},\boldsymbol{\alpha}_{1}\right)\right\rangle\underset{(b)}{=}\\
&\left\langle\boldsymbol{x}_{2}^{*}-\boldsymbol{x}_{1}^{*},F\left(\boldsymbol{x}_{2}^{*},\boldsymbol{\alpha}_{2}\right)-F\left(\boldsymbol{x}_{1}^{*},\boldsymbol{\alpha}_{1}\right)\right\rangle+\\
&\left\langle\sum_{n=1}^{N}\left(\boldsymbol{x}_{n}^{*}\left(\boldsymbol{\alpha}_{2}\right)-\boldsymbol{x}_{n}^{*}\left(\boldsymbol{\alpha}_{1}\right)\right),\boldsymbol{\alpha}_{2}-\boldsymbol{\alpha}_{1}\right\rangle\underset{(c)}{\geq}\\
&\left\langle\sum_{n=1}^{N}\left(\boldsymbol{x}_{n}^{*}\left(\boldsymbol{\alpha}_{2}\right)-\boldsymbol{x}_{n}^{*}\left(\boldsymbol{\alpha}_{1}\right)\right),\boldsymbol{\alpha}_{2}-\boldsymbol{\alpha}_{1}\right\rangle
\end{aligned} \tag{41}$$

where (a) uses that $F\left(\boldsymbol{x},\boldsymbol{\alpha}_{1}\right)$ is strongly monotone in $\boldsymbol{x}$ with parameter $\lambda>0$, (b) uses the linearity of $F\left(\boldsymbol{x},\boldsymbol{\alpha}\right)$ in $\boldsymbol{\alpha}$ and (c) uses (40).

Now let $\boldsymbol{\alpha}_{1},\boldsymbol{\alpha}_{2}\in\mathcal{A}^{*}$ and let $\boldsymbol{x}^{*}\left(\boldsymbol{\alpha}_{1}\right),\boldsymbol{x}^{*}\left(\boldsymbol{\alpha}_{2}\right)$ be the corresponding NE. For every k,

- If $\alpha_{1}^{k}=\alpha_{2}^{k}=0$ then $\left(\alpha_{2}^{k}-\alpha_{1}^{k}\right)\sum_{n=1}^{N}\left(x_{n}^{*k}\left(\boldsymbol{\alpha}_{2}\right)-x_{n}^{*k}\left(\boldsymbol{\alpha}_{1}\right)\right)=0$.
- If $\alpha_{1}^{k}>0$ and $\alpha_{2}^{k}>0$ then $\sum_{n=1}^{N}x_{n}^{*k}\left(\boldsymbol{\alpha}_{1}\right)=\sum_{n=1}^{N}x_{n}^{*k}\left(\boldsymbol{\alpha}_{2}\right)=l_{k}^{*}$ so $\left(\alpha_{2}^{k}-\alpha_{1}^{k}\right)\sum_{n=1}^{N}\left(x_{n}^{*k}\left(\boldsymbol{\alpha}_{2}\right)-x_{n}^{*k}\left(\boldsymbol{\alpha}_{1}\right)\right)=0$.
- If $\alpha_{1}^{k}>0$ and $\alpha_{2}^{k}=0$ then $\sum_{n=1}^{N}x_{n}^{*k}\left(\boldsymbol{\alpha}_{2}\right)<l_{k}^{*}=\sum_{n=1}^{N}x_{n}^{*k}\left(\boldsymbol{\alpha}_{1}\right)$ so $\left(\alpha_{2}^{k}-\alpha_{1}^{k}\right)\sum_{n=1}^{N}\left(x_{n}^{*k}\left(\boldsymbol{\alpha}_{2}\right)-x_{n}^{*k}\left(\boldsymbol{\alpha}_{1}\right)\right)\geq 0$.
- If $\alpha_{1}^{k}=0$ and $\alpha_{2}^{k}>0$ then $\sum_{n=1}^{N}x_{n}^{*k}\left(\boldsymbol{\alpha}_{2}\right)=l_{k}^{*}>\sum_{n=1}^{N}x_{n}^{*k}\left(\boldsymbol{\alpha}_{1}\right)$ so $\left(\alpha_{2}^{k}-\alpha_{1}^{k}\right)\sum_{n=1}^{N}\left(x_{n}^{*k}\left(\boldsymbol{\alpha}_{2}\right)-x_{n}^{*k}\left(\boldsymbol{\alpha}_{1}\right)\right)\geq 0$.

We conclude that if $\boldsymbol{\alpha}_{1},\boldsymbol{\alpha}_{2}\in\mathcal{A}^{*}$ then

$$\left\langle\sum_{n=1}^{N}\left(\boldsymbol{x}_{n}^{*}\left(\boldsymbol{\alpha}_{2}\right)-\boldsymbol{x}_{n}^{*}\left(\boldsymbol{\alpha}_{1}\right)\right),\boldsymbol{\alpha}_{2}-\boldsymbol{\alpha}_{1}\right\rangle\geq 0 \tag{42}$$

which by (41) implies that $\boldsymbol{x}^{*}\left(\boldsymbol{\alpha}_{2}\right)=\boldsymbol{x}^{*}\left(\boldsymbol{\alpha}_{1}\right)$.

13. Proof of Lemma 5

First we bound the distance between $\boldsymbol{x}_t$ and the new NE $\boldsymbol{x}^*(\boldsymbol{\alpha}_t)$. With probability 1, for some constant $C_0 > 0$,

$$\begin{aligned}
&\left\|\boldsymbol{x}_t - \boldsymbol{x}^*(\boldsymbol{\alpha}_t)\right\|^2 = \\
&\left\|\boldsymbol{x}_t - \boldsymbol{x}^*(\boldsymbol{\alpha}_{t-1}) + \boldsymbol{x}^*(\boldsymbol{\alpha}_{t-1}) - \boldsymbol{x}^*(\boldsymbol{\alpha}_t)\right\|^2 \underset{(a)}{\leq} \\
&\left\|\boldsymbol{x}_t - \boldsymbol{x}^*(\boldsymbol{\alpha}_{t-1})\right\|^2 + \left\|\boldsymbol{x}^*(\boldsymbol{\alpha}_t) - \boldsymbol{x}^*(\boldsymbol{\alpha}_{t-1})\right\|^2 + \\
&2\left\|\boldsymbol{x}_t - \boldsymbol{x}^*(\boldsymbol{\alpha}_{t-1})\right\|\left\|\boldsymbol{x}^*(\boldsymbol{\alpha}_t) - \boldsymbol{x}^*(\boldsymbol{\alpha}_{t-1})\right\| \underset{(b)}{\leq} \\
&\left\|\boldsymbol{x}_t - \boldsymbol{x}^*(\boldsymbol{\alpha}_{t-1})\right\|^2 + \varepsilon_{t-1}^2 L^2 \left\|\sum_{n=1}^N \boldsymbol{x}_{n,t} - \boldsymbol{l}^*\right\|^2 \\
&+ 2\varepsilon_{t-1} L \left\|\sum_{n=1}^N \boldsymbol{x}_{n,t} - \boldsymbol{l}^*\right\| \left\|\boldsymbol{x}_t - \boldsymbol{x}^*(\boldsymbol{\alpha}_{t-1})\right\| \underset{(c)}{\leq} \\
&\left\|\boldsymbol{x}_t - \boldsymbol{x}^*(\boldsymbol{\alpha}_{t-1})\right\|^2 + C_0\varepsilon_t
\end{aligned} \tag{43}$$

where (a) is Cauchy-Schwarz, (b) follows from Lemma 3 and (c) uses that $\boldsymbol{x}_{n,t}, \boldsymbol{x}_t$ and $\boldsymbol{x}^*(\boldsymbol{\alpha}_{t-1})$ are bounded and that $\frac{\varepsilon_{t+1}}{\varepsilon_t} \to 1$ as $t \to \infty$ (condition 4 of Theorem 1).

Note that $\mathbf{1}_N \otimes \boldsymbol{\alpha}_t$ concatenates $\boldsymbol{\alpha}_t$ N times. Next we bound the norm of the stochastic gradient vector. With probability 1, we have that for some constants $B_0, B_1, B_2 > 0$,

$$\begin{aligned}
&\eta_t \left\|\boldsymbol{g}_t - \mathbf{1}_N \otimes \boldsymbol{\alpha}_t\right\| \leq \eta_t\left(\sqrt{N}\left\|\boldsymbol{\alpha}_t\right\| + \left\|\boldsymbol{g}_t\right\|\right) \underset{(a)}{\leq} \\
&\eta_t B_0 \left(\left\|\boldsymbol{\alpha}_0\right\| + \sum_{\tau=0}^{t-1}\varepsilon_\tau \left\|\sum_{n=1}^N \boldsymbol{x}_{n,\tau+1} - \boldsymbol{l}^*\right\| + \left\|\boldsymbol{g}_t\right\|\right) \leq \\
&\eta_t B_1 \left(\sum_{\tau=0}^{t-1}\varepsilon_\tau + \left\|\boldsymbol{g}_t\right\|\right) \underset{(b)}{\leq} B_2\left(\sqrt{\varepsilon_t} + \eta_t\left\|\boldsymbol{g}_t\right\|\right)
\end{aligned} \tag{44}$$

where (a) iterates over

$$\begin{aligned}
\left\|\boldsymbol{\alpha}_{t+1}\right\| &= \left\|\left[\boldsymbol{\alpha}_t + \varepsilon_t\left(\sum_{n=1}^N \boldsymbol{x}_{n,t+1} - \boldsymbol{l}^*\right)\right]^+\right\| \leq \\
&\left\|\boldsymbol{\alpha}_t\right\| + \varepsilon_t \left\|\sum_{n=1}^N \boldsymbol{x}_{n,t+1} - \boldsymbol{l}^*\right\|
\end{aligned} \tag{45}$$

and (b) uses condition 3 of Theorem 1. To see that, let $\rho > 0$. Then pick a large enough T_0 such that for all $t > T_0$ we have $B_1 \frac{\eta_t \sum_{\tau=0}^{t-1}\varepsilon_\tau}{\sqrt{\varepsilon_t}} < \rho$. Hence we can use $B_2 = \max\left\{B_1 \max_{0 \leq t \leq T_0} \frac{\eta_t \sum_{\tau=0}^{t-1}\varepsilon_\tau}{\sqrt{\varepsilon_t}}, \rho, B_1\right\}$.

Now we can analyze the gradient behavior. Recall the definition of $F(\boldsymbol{x}, \boldsymbol{\alpha})$ in (16). Then, with probability 1, for some constants $C_1, C_2, C_3 > 0$,

$$\begin{aligned}
&\eta_t \mathbb{E}\left\{\left\langle \boldsymbol{x}_t - \boldsymbol{x}^*(\boldsymbol{\alpha}_t), \boldsymbol{g}_t - \mathbf{1}_N \otimes \boldsymbol{\alpha}_t \right\rangle \mid \mathcal{F}_t\right\} = \\
&\eta_t \mathbb{E}\left\{\left\langle \boldsymbol{x}_t - \boldsymbol{x}^*(\boldsymbol{\alpha}_{t-1}), \boldsymbol{g}_t - \mathbf{1}_N \otimes \boldsymbol{\alpha}_t \right\rangle \mid \mathcal{F}_t\right\} \\
&+ \eta_t \mathbb{E}\left\{\left\langle \boldsymbol{x}^*(\boldsymbol{\alpha}_{t-1}) - \boldsymbol{x}^*(\boldsymbol{\alpha}_t), \boldsymbol{g}_t - \mathbf{1}_N \otimes \boldsymbol{\alpha}_t \right\rangle \mid \mathcal{F}_t\right\} \leq \\
&\eta_t \left\langle \boldsymbol{x}_t - \boldsymbol{x}^*(\boldsymbol{\alpha}_{t-1}), \mathbb{E}\left\{\boldsymbol{g}_t - \mathbf{1}_N \otimes \boldsymbol{\alpha}_t | \mathcal{F}_t\right\} - F(\boldsymbol{x}_t, \boldsymbol{\alpha}_{t-1})\right\rangle \\
&+ \eta_t \left\langle \boldsymbol{x}_t - \boldsymbol{x}^*(\boldsymbol{\alpha}_{t-1}), F(\boldsymbol{x}_t, \boldsymbol{\alpha}_{t-1})\right\rangle \\
&+ \eta_t \mathbb{E}\left\{\left\|\boldsymbol{x}^*(\boldsymbol{\alpha}_t) - \boldsymbol{x}^*(\boldsymbol{\alpha}_{t-1})\right\| \left\|\boldsymbol{g}_t - \mathbf{1}_N \otimes \boldsymbol{\alpha}_t\right\| \mid \mathcal{F}_t\right\} \underset{(a)}{\leq} \\
&\sqrt{N}\eta_t \left\|\boldsymbol{x}_t - \boldsymbol{x}^*(\boldsymbol{\alpha}_{t-1})\right\| \left\|\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{t-1}\right\| \\
&+ \eta_t \left\|\boldsymbol{x}_t - \boldsymbol{x}^*(\boldsymbol{\alpha}_{t-1})\right\| \left\|\mathbb{E}\left\{\boldsymbol{g}_t \mid \mathcal{F}_t\right\} - F(\boldsymbol{x}_t)\right\| \\
&- \lambda\eta_t \left\|\boldsymbol{x}_t - \boldsymbol{x}^*(\boldsymbol{\alpha}_{t-1})\right\|^2 \\
&+ C_1 \varepsilon_t \eta_t \mathbb{E}\left\{\left\|\boldsymbol{g}_t - \mathbf{1}_N \otimes \boldsymbol{\alpha}_t\right\| \mid \mathcal{F}_t\right\} \underset{(b)}{\leq} \\
&\eta_t \delta_t \left\|\boldsymbol{x}_t - \boldsymbol{x}^*(\boldsymbol{\alpha}_{t-1})\right\| - \lambda\eta_t \left\|\boldsymbol{x}_t - \boldsymbol{x}^*(\boldsymbol{\alpha}_{t-1})\right\|^2 \\
&+ C_2 \varepsilon_t^{3/2} + C_3 \eta_t \varepsilon_t
\end{aligned} \tag{46}$$

where (a) uses that $\left\|\boldsymbol{x}^*(\boldsymbol{\alpha}_t) - \boldsymbol{x}^*(\boldsymbol{\alpha}_{t-1})\right\| \leq C_1\varepsilon_t$ (Lemma 3 and $\frac{\varepsilon_{t+1}}{\varepsilon_t} \to 1$), and also that since $F(\boldsymbol{x}, \boldsymbol{\alpha}_{t-1})$ is strongly monotone in $\boldsymbol{x}$ with parameter $\lambda > 0$, then

$$\begin{aligned}
&\left\langle \boldsymbol{x}_t - \boldsymbol{x}^*(\boldsymbol{\alpha}_{t-1}), F(\boldsymbol{x}_t, \boldsymbol{\alpha}_{t-1})\right\rangle \underset{(a)}{\leq} \\
&\left\langle \boldsymbol{x}_t - \boldsymbol{x}^*(\boldsymbol{\alpha}_{t-1}), F(\boldsymbol{x}_t, \boldsymbol{\alpha}_{t-1}) - F(\boldsymbol{x}^*(\boldsymbol{\alpha}_{t-1}), \boldsymbol{\alpha}_{t-1})\right\rangle \\
&\leq -\lambda \left\|\boldsymbol{x}_t - \boldsymbol{x}^*(\boldsymbol{\alpha}_{t-1})\right\|^2
\end{aligned} \tag{47}$$

where (a) follows since $\left\langle \boldsymbol{x} - \boldsymbol{x}^*(\boldsymbol{\alpha}), F(\boldsymbol{x}^*(\boldsymbol{\alpha}), \boldsymbol{\alpha})\right\rangle \leq 0$ for all $\boldsymbol{\alpha} \in \mathbb{R}_+^K$ and $\boldsymbol{x} \in \mathcal{X}$, since $\boldsymbol{x}^*(\boldsymbol{\alpha})$ is a NE (see Proposition 1.4.2 in (Facchinei & Pang, 2007)). Inequality (b) in (46) follows from (44) and the assumption in Definition 3.

Now we can bound how the distance from NE evolves. Then, with probability 1, for some constants $C_4, C_5, C_6 > 0$,

$$\begin{aligned}
&\mathbb{E}\left\{\left\|\boldsymbol{x}_{t+1} - \boldsymbol{x}^*(\boldsymbol{\alpha}_t)\right\|^2 \mid \mathcal{F}_t\right\} \underset{(a)}{\leq} \\
&\mathbb{E}\left\{\left\|\boldsymbol{x}_t + \eta_t(\boldsymbol{g}_t - \mathbf{1}_N \otimes \boldsymbol{\alpha}_t) - \boldsymbol{x}^*(\boldsymbol{\alpha}_t)\right\|^2 \mid \mathcal{F}_t\right\} = \\
&\mathbb{E}\left\{\left\|\boldsymbol{x}_t - \boldsymbol{x}^*(\boldsymbol{\alpha}_t)\right\|^2 + \eta_t^2 \left\|\boldsymbol{g}_t - \mathbf{1}_N \otimes \boldsymbol{\alpha}_t\right\|^2 \mid \mathcal{F}_t\right\} \\
&+ 2\eta_t \mathbb{E}\left\{\left\langle \boldsymbol{x}_t - \boldsymbol{x}^*(\boldsymbol{\alpha}_t), \boldsymbol{g}_t - \mathbf{1}_N \otimes \boldsymbol{\alpha}_t\right\rangle \mid \mathcal{F}_t\right\} \underset{(b)}{\leq} \\
&\left\|\boldsymbol{x}_t - \boldsymbol{x}^*(\boldsymbol{\alpha}_{t-1})\right\|^2 + C_0\varepsilon_t + C_4\left(\varepsilon_t + \eta_t^2 \mathbb{E}\left\{\left\|\boldsymbol{g}_t\right\|^2 \mid \mathcal{F}_t\right\}\right) + \\
&2\eta_t\delta_t \left\|\boldsymbol{x}_t - \boldsymbol{x}^*(\boldsymbol{\alpha}_{t-1})\right\| - 2\lambda\eta_t \left\|\boldsymbol{x}_t - \boldsymbol{x}^*(\boldsymbol{\alpha}_{t-1})\right\|^2 \\
&+ 2C_2\varepsilon_t^{3/2} + 2C_3\eta_t\varepsilon_t \underset{(c)}{\leq} \\
&(1 - 2\eta_t(\lambda - \delta_t))\left\|\boldsymbol{x}_t - \boldsymbol{x}^*(\boldsymbol{\alpha}_{t-1})\right\|^2 + 2\eta_t\delta_t + C_5\varepsilon_t + C_6\eta_t^2
\end{aligned} \tag{48}$$

where (a) uses $\|\Pi_{\mathcal{X}}\boldsymbol{y}-\boldsymbol{x}\| \leq \|\boldsymbol{y}-\boldsymbol{x}\|$ for any $\boldsymbol{x} \in \mathcal{X}$ since $\mathcal{X}$ is convex. Inequality (b) uses (43), (44), and (46) and inequality (c) uses Definition 3 and $x \leq x^2+1$.

The last step of the proof is to use (48) to show by induction that for every $t \geq 1$

$$\mathbb{E}\left\{\|\boldsymbol{x}_t-\boldsymbol{x}^*(\boldsymbol{\alpha}_{t-1})\|^2\right\} \leq A\frac{\varepsilon_t}{\eta_t} \tag{49}$$

for some $A>0$. First we define T_0 to be large enough such that $\delta_t \leq \frac{\lambda}{4}$, $\max\left\{\eta_t^2,\eta_t\delta_t\right\} \leq C_7\varepsilon_t$ for some $C_7>0$ and also that $\frac{\varepsilon_t-\varepsilon_{t+1}}{\eta_t} \leq \lambda\varepsilon_t$ for all $t>T_0$ (using conditions 1,2,4 of Theorem 1). Then we pick $A=\max\left\{\max\limits_{1\leq t\leq T_0}\frac{\eta_t}{\varepsilon_t}\mathbb{E}\left\{\|\boldsymbol{x}_t-\boldsymbol{x}^*(\boldsymbol{\alpha}_{t-1})\|^2\right\},A_0\right\}$ for some A_0 that is specified below, which is a constant with respect to t. Hence for all $1\leq t\leq T_0$ (49) holds. For $t>T_0$ we take the expectation on both sides of (48) to obtain

$$\begin{aligned}&\mathbb{E}\left\{\|\boldsymbol{x}_{t+1}-\boldsymbol{x}^*(\boldsymbol{\alpha}_t)\|^2\right\} \leq \\ &\quad(1-2\eta_t(\lambda-\delta_t))\,\mathbb{E}\left\{\|\boldsymbol{x}_t-\boldsymbol{x}^*(\boldsymbol{\alpha}_{t-1})\|^2\right\}\\ &+2\eta_t\delta_t+C_5\varepsilon_t+C_6\eta_t^2 \underset{(a)}{\leq} \left(1-\frac{3}{2}\eta_t\lambda\right)A\frac{\varepsilon_t}{\eta_t}+D_0\varepsilon_t=\\ &A\frac{\varepsilon_t}{\eta_t}+\left(D_0-\frac{3}{2}\lambda A\right)\varepsilon_t \underset{(b)}{\leq} A\left(\frac{\varepsilon_t}{\eta_t}-\lambda\varepsilon_t\right) \underset{(c)}{\leq} A\frac{\varepsilon_{t+1}}{\eta_{t+1}}\end{aligned} \tag{50}$$

where (a) follows for some constant $D_0>0$ since $\delta_t \leq \frac{\lambda}{4}$ and $\max\left\{\eta_t^2,\eta_t\delta_t\right\} \leq C_7\varepsilon_t$ for $t>T_0$. In (b) we used $A \geq \frac{2D_0}{\lambda}$ so we set $A_0=\frac{2D_0}{\lambda}$ and in (c) we used that $\frac{\varepsilon_t-\varepsilon_{t+1}}{\eta_t} \leq \lambda\varepsilon_t$ so $\frac{\varepsilon_{t+1}}{\eta_{t+1}} \geq \frac{\varepsilon_{t+1}}{\eta_t} \geq \frac{\varepsilon_t}{\eta_t}-\lambda\varepsilon_t$.

14. Proof of Theorem 1

We have that with probability 1

$$\begin{aligned}&\min_{\boldsymbol{\alpha}^*\in\mathcal{A}^*}\|\boldsymbol{\alpha}_{t+1}-\boldsymbol{\alpha}^*\|^2=\\ &\min_{\boldsymbol{\alpha}^*\in\mathcal{A}^*}\left\|\left[\boldsymbol{\alpha}_t+\varepsilon_t\left(\sum_{n=1}^N\boldsymbol{x}_{n,t+1}-\boldsymbol{l}^*\right)\right]^+-\boldsymbol{\alpha}^*\right\|^2 \underset{(a)}{\leq}\\ &\min_{\boldsymbol{\alpha}^*\in\mathcal{A}^*}\left\|\boldsymbol{\alpha}_t+\varepsilon_t\left(\sum_{n=1}^N\boldsymbol{x}_{n,t+1}-\boldsymbol{l}^*\right)-\boldsymbol{\alpha}^*\right\|^2=\\ &\min_{\boldsymbol{\alpha}^*\in\mathcal{A}^*}\left[\|\boldsymbol{\alpha}_t-\boldsymbol{\alpha}^*\|^2+2\varepsilon_t\left\langle\sum_{n=1}^N\boldsymbol{x}_{n,t+1}-\boldsymbol{l}^*,\boldsymbol{\alpha}_t-\boldsymbol{\alpha}^*\right\rangle\right]\\ &+\varepsilon_t^2\left\|\sum_{n=1}^N\boldsymbol{x}_{n,t+1}-\boldsymbol{l}^*\right\|^2 \underset{(b)}{\leq}\\ &\min_{\boldsymbol{\alpha}^*\in\mathcal{A}^*}\left[\|\boldsymbol{\alpha}_t-\boldsymbol{\alpha}^*\|^2+2\varepsilon_t\left\langle\sum_{n=1}^N\left(\boldsymbol{x}_{n,t+1}-\boldsymbol{x}_n^*(\boldsymbol{\alpha}_t)\right),\boldsymbol{\alpha}_t-\boldsymbol{\alpha}^*\right\rangle\right.\\ &+2\varepsilon_t\left\langle\sum_{n=1}^N\boldsymbol{x}_n^*(\boldsymbol{\alpha}_t)-\boldsymbol{x}_n^*(\boldsymbol{\alpha}^*),\boldsymbol{\alpha}_t-\boldsymbol{\alpha}^*\right\rangle\\ &\left.+2\varepsilon_t\left\langle\sum_{n=1}^N\boldsymbol{x}_n^*(\boldsymbol{\alpha}^*)-\boldsymbol{l}^*,\boldsymbol{\alpha}_t-\boldsymbol{\alpha}^*\right\rangle\right]+\varepsilon_t^2D_1 \underset{(c)}{\leq}\\ &\min_{\boldsymbol{\alpha}^*\in\mathcal{A}^*}\left[\|\boldsymbol{\alpha}_t-\boldsymbol{\alpha}^*\|^2+2\varepsilon_t\left\|\sum_{n=1}^N\left(\boldsymbol{x}_{n,t+1}-\boldsymbol{x}_n^*(\boldsymbol{\alpha}_t)\right)\right\|\|\boldsymbol{\alpha}_t-\boldsymbol{\alpha}^*\|\right.\\ &+2\varepsilon_t\left\langle\sum_{n=1}^N\left(\boldsymbol{x}_n^*(\boldsymbol{\alpha}_t)-\boldsymbol{x}_n^*(\boldsymbol{\alpha}^*)\right),\boldsymbol{\alpha}_t-\boldsymbol{\alpha}^*\right\rangle\\ &\left.+2\varepsilon_t\left\langle\sum_{n=1}^N\boldsymbol{x}_n^*(\boldsymbol{\alpha}^*)-\boldsymbol{l}^*,\boldsymbol{\alpha}_t-\boldsymbol{\alpha}^*\right\rangle\right]+\varepsilon_t^2D_1 \underset{(d)}{\leq}\\ &\left(1+2\varepsilon_t\left\|\sum_{n=1}^N\left(\boldsymbol{x}_{n,t+1}-\boldsymbol{x}_n^*(\boldsymbol{\alpha}_t)\right)\right\|\right)\min_{\boldsymbol{\alpha}^*\in\mathcal{A}^*}\|\boldsymbol{\alpha}_t-\boldsymbol{\alpha}^*\|^2\\ &+2\varepsilon_t\max_{\boldsymbol{\alpha}^*\in\mathcal{A}^*}\left[\left\langle\sum_{n=1}^N\left(\boldsymbol{x}_n^*(\boldsymbol{\alpha}_t)-\boldsymbol{x}_n^*(\boldsymbol{\alpha}^*)\right),\boldsymbol{\alpha}_t-\boldsymbol{\alpha}^*\right\rangle\right.\\ &\left.+\left\langle\sum_{n=1}^N\boldsymbol{x}_n^*(\boldsymbol{\alpha}^*)-\boldsymbol{l}^*,\boldsymbol{\alpha}_t-\boldsymbol{\alpha}^*\right\rangle\right]\\ &+2\varepsilon_t\left\|\sum_{n=1}^N\left(\boldsymbol{x}_{n,t+1}-\boldsymbol{x}_n^*(\boldsymbol{\alpha}_t)\right)\right\|+\varepsilon_t^2D_1\end{aligned} \tag{51}$$

where (a) follows since $[\boldsymbol{x}]^+$ can only decrease the distance of $\boldsymbol{x}$ to the set $\mathcal{A}^*$ since $\boldsymbol{\alpha}^* \geq 0$ for all $\boldsymbol{\alpha}^* \in \mathcal{A}^*$. Inequality (b) uses that $\left\|\sum_{n=1}^N\boldsymbol{x}_{n,t+1}-\boldsymbol{l}^*\right\|^2 \leq D_1$ for some $D_1>0$ since $\mathcal{X}$ is bounded, (c) is Cauchy-Schwarz and (d) uses that $\min\limits_{\boldsymbol{x}}\left(f(\boldsymbol{x})+g(\boldsymbol{x})\right) \leq \min\limits_{\boldsymbol{x}}f(\boldsymbol{x})+\max\limits_{\boldsymbol{x}}g(\boldsymbol{x})$ for any functions $f(\boldsymbol{x}),g(\boldsymbol{x})$ and then uses $\|\boldsymbol{\alpha}_t-\boldsymbol{\alpha}^*\| \leq$

$\|\boldsymbol{\alpha}_t - \boldsymbol{\alpha}^*\|^2 + 1$. Define

$$f(\boldsymbol{\alpha}) = -\max_{\boldsymbol{\alpha}^* \in \mathcal{A}^*} \left[\left\langle \sum_{n=1}^{N} \left(\boldsymbol{x}_n^*(\boldsymbol{\alpha}) - \boldsymbol{x}_n^*(\boldsymbol{\alpha}^*)\right), \boldsymbol{\alpha} - \boldsymbol{\alpha}^* \right\rangle + \left\langle \sum_{n=1}^{N} \boldsymbol{x}_n^*(\boldsymbol{\alpha}^*) - \boldsymbol{l}^*, \boldsymbol{\alpha} - \boldsymbol{\alpha}^* \right\rangle \right]. \quad (52)$$

Then $\left\langle \sum_{n=1}^{N} \left(\boldsymbol{x}_n^*(\boldsymbol{\alpha}) - \boldsymbol{x}_n^*(\boldsymbol{\alpha}^*)\right), \boldsymbol{\alpha} - \boldsymbol{\alpha}^* \right\rangle \leq 0$ by Lemma 4 and $\left\langle \sum_{n=1}^{N} \boldsymbol{x}_n^*(\boldsymbol{\alpha}^*) - \boldsymbol{l}^*, \boldsymbol{\alpha} - \boldsymbol{\alpha}^* \right\rangle \leq 0$ by Lemma 1, for any $\boldsymbol{\alpha}^* \in \mathcal{A}^*$, so $f(\boldsymbol{\alpha}) \geq 0$ for all $\boldsymbol{\alpha} \in \mathbb{R}_+^K$.

Define the filtration $\mathcal{G}_t \triangleq \mathcal{F}_{t+1}$. By taking the conditional expectation on both sides of (51):

$$\begin{aligned} &\mathbb{E}\left\{ \min_{\boldsymbol{\alpha}^* \in \mathcal{A}^*} \|\boldsymbol{\alpha}_{t+1} - \boldsymbol{\alpha}^*\|^2 \mid \mathcal{G}_t \right\} \leq \\ &\mathbb{E}\left\{ (1 + 2N\varepsilon_t \|\boldsymbol{x}^*(\boldsymbol{\alpha}_t) - \boldsymbol{x}_{t+1}\|) \min_{\boldsymbol{\alpha}^* \in \mathcal{A}^*} \|\boldsymbol{\alpha}_t - \boldsymbol{\alpha}^*\|^2 \mid \mathcal{G}_t \right\} \\ &- 2\varepsilon_t f(\boldsymbol{\alpha}_t) + 2N\varepsilon_t \mathbb{E}\left\{ \|\boldsymbol{x}^*(\boldsymbol{\alpha}_t) - \boldsymbol{x}_{t+1}\| \mid \mathcal{G}_t \right\} + \varepsilon_t^2 D_1 = \\ &(1 + 2N\varepsilon_t \|\boldsymbol{x}^*(\boldsymbol{\alpha}_t) - \boldsymbol{x}_{t+1}\|) \min_{\boldsymbol{\alpha}^* \in \mathcal{A}^*} \|\boldsymbol{\alpha}_t - \boldsymbol{\alpha}^*\|^2 \\ &- 2\varepsilon_t f(\boldsymbol{\alpha}_t) + 2N\varepsilon_t \|\boldsymbol{x}^*(\boldsymbol{\alpha}_t) - \boldsymbol{x}_{t+1}\| + \varepsilon_t^2 D_1. \end{aligned} \quad (53)$$

Next we want to apply Robbins-Siegmund "almost" super-martingale Theorem (Robbins & Siegmund, 1971) based on (53). Using their notation, we have $z_t = \min_{\boldsymbol{\alpha}^* \in \mathcal{A}^*} \|\boldsymbol{\alpha}_t - \boldsymbol{\alpha}^*\|^2$, $\beta_t = 2N\varepsilon_t \|\boldsymbol{x}^*(\boldsymbol{\alpha}_t) - \boldsymbol{x}_{t+1}\|$, $\xi_t = 2N\varepsilon_t \|\boldsymbol{x}^*(\boldsymbol{\alpha}_t) - \boldsymbol{x}_{t+1}\| + \varepsilon_t^2 D_1$ and $\zeta_t = 2\varepsilon_t f(\boldsymbol{\alpha}_t) \geq 0$.

From Lemma 5 we obtain for some $A > 0$ that

$$\mathbb{E}\left\{ \sum_{t=0}^{\infty} \varepsilon_t \|\boldsymbol{x}^*(\boldsymbol{\alpha}_t) - \boldsymbol{x}_{t+1}\| \right\} \underset{(a)}{\leq} \sum_{t=0}^{\infty} \varepsilon_t \mathbb{E}\left\{ \|\boldsymbol{x}^*(\boldsymbol{\alpha}_t) - \boldsymbol{x}_{t+1}\| \right\} \leq A \sum_{t=0}^{\infty} \varepsilon_t \sqrt{\frac{\varepsilon_t}{\eta_t}} \underset{(b)}{<} \infty \quad (54)$$

where (a) is Fatou's Lemma (Billingsley, 2008) and (b) is condition 3 of the Theorem. Hence, $\sum_{t=0}^{\infty} \varepsilon_t \|\boldsymbol{x}^*(\boldsymbol{\alpha}_t) - \boldsymbol{x}_{t+1}\| < \infty$ with probability 1 so $\sum_{t=0}^{\infty} \beta_t < \infty$ and $\sum_{t=0}^{\infty} \xi_t < \infty$, also using that $\sum_{t=0}^{\infty} \varepsilon_t^2 < \infty$. Then (Robbins & Siegmund, 1971) states that with probability 1

$$\lim_{t \to \infty} \min_{\boldsymbol{\alpha}^* \in \mathcal{A}^*} \|\boldsymbol{\alpha}_t - \boldsymbol{\alpha}^*\|^2 = \Lambda \quad (55)$$

for some random variable Λ, and that with probability 1

$$\sum_{t=0}^{\infty} \varepsilon_t f(\boldsymbol{\alpha}_t) < \infty. \quad (56)$$

Now we show that for every $\boldsymbol{\alpha} \notin \mathcal{A}^*$ we have $f(\boldsymbol{\alpha}) > 0$. If $f(\tilde{\boldsymbol{\alpha}}) = 0$ then both the non-negative terms in (52) are zero, so Lemma 4 implies that if $f(\tilde{\boldsymbol{\alpha}}) = 0$ then there must exists an $\boldsymbol{\alpha}^* \in \mathcal{A}^*$ such that $\boldsymbol{x}^*(\tilde{\boldsymbol{\alpha}}) = \boldsymbol{x}^*(\boldsymbol{\alpha}^*)$ and

$$0 = \left\langle \sum_{n=1}^{N} \boldsymbol{x}_n^*(\boldsymbol{\alpha}^*) - \boldsymbol{l}^*, \tilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^* \right\rangle \underset{(a)}{=} \left\langle \sum_{n=1}^{N} \boldsymbol{x}_n^*(\boldsymbol{\alpha}^*) - \boldsymbol{l}^*, \tilde{\boldsymbol{\alpha}} \right\rangle = \left\langle \sum_{n=1}^{N} \boldsymbol{x}_n^*(\tilde{\boldsymbol{\alpha}}) - \boldsymbol{l}^*, \tilde{\boldsymbol{\alpha}} \right\rangle \quad (57)$$

where (a) uses $\left\langle \sum_{n=1}^{N} \boldsymbol{x}_n^*(\boldsymbol{\alpha}^*) - \boldsymbol{l}^*, \boldsymbol{\alpha}^* \right\rangle = 0$ (Lemma 1). Hence, by invoking Lemma 1 again, (57) and $\sum_{n=1}^{N} \boldsymbol{x}_n^*(\tilde{\boldsymbol{\alpha}}) = \sum_{n=1}^{N} \boldsymbol{x}_n^*(\boldsymbol{\alpha}^*) \leq \boldsymbol{l}^*$ imply that $(\boldsymbol{x}^*(\tilde{\boldsymbol{\alpha}}), \tilde{\boldsymbol{\alpha}})$ solves the VI in (17), so $\tilde{\boldsymbol{\alpha}} \in \mathcal{A}^*$.

Let $\omega \in \Omega$ such that $\Lambda(\omega) > 0$, if exists. Then there exist $T_0(\omega)$ and $a(\omega), b(\omega) > 0$ such that for all $t \geq T_0(\omega)$

$$\boldsymbol{\alpha}_t \in \mathcal{R}(\omega) = \left\{ \boldsymbol{\alpha} \,\middle|\, a(\omega) \leq \min_{\boldsymbol{\alpha}^* \in \mathcal{A}^*} \|\boldsymbol{\alpha} - \boldsymbol{\alpha}^*\|^2 \leq b(\omega) \right\}. \quad (58)$$

Since $\mathcal{A}^*$ is compact (Lemma 2 shows that $\mathcal{X}^* \times \mathcal{A}^*$ is compact) then $g(\boldsymbol{\alpha}) = \min_{\boldsymbol{\alpha}^* \in \mathcal{A}^*} \|\boldsymbol{\alpha} - \boldsymbol{\alpha}^*\|^2$ is continuous, which makes $\mathcal{R}(\omega)$ compact as well.

Since $\boldsymbol{x}^*(\boldsymbol{\alpha})$ is continuous (Lemma 3), then $f(\boldsymbol{\alpha})$ is continuous as a maximum of continuous functions over the compact $\mathcal{A}^*$ (Lemma 2). Hence, $\min_{\boldsymbol{\alpha} \in \mathcal{R}(\omega)} f(\boldsymbol{\alpha}) = c$ for some $c > 0$. But then $\boldsymbol{\alpha}_t \in \mathcal{R}(\omega)$ for all $t \geq T_0(\omega)$ implies that $\sum_{t=0}^{\infty} \varepsilon_t f(\boldsymbol{\alpha}_t) = \infty$, which by (56) cannot occur for more than a measure zero set of $\omega \in \Omega$. We conclude that $\min_{\boldsymbol{\alpha}^* \in \mathcal{A}^*} \|\boldsymbol{\alpha}_t - \boldsymbol{\alpha}^*\|^2 \to 0$ as $t \to \infty$ with probability 1.

Let $\varepsilon > 0$. The above implies that for almost all $\omega \in \Omega$ there exists $T(\omega)$ such that for all $t > T(\omega)$ there is a sequence $\boldsymbol{\alpha}_t^*(\omega) \in \mathcal{A}^*$ such that

$$\|\boldsymbol{x}^*(\boldsymbol{\alpha}_t) - \boldsymbol{x}^*(\boldsymbol{\alpha}_t^*(\omega))\| \underset{(a)}{\leq} L \|\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_t^*(\omega)\| \leq \varepsilon \quad (59)$$

where (a) follows from the Lipschitz continuity of $\boldsymbol{x}^*(\boldsymbol{\alpha})$ (Lemma 3). However, by Lemma 4 we know that $\boldsymbol{x}^*(\boldsymbol{\alpha}) = \boldsymbol{x}^*$ for all $\boldsymbol{\alpha} \in \mathcal{A}^*$ for some NE $\boldsymbol{x}^*$ such that for all k

$$\sum_{n=1}^{N} x_n^{*k} = l_k^* \text{ or } \left[\sum_{n=1}^{N} x_n^{*k} < l_k^* \text{ and } \alpha_k = 0 \right]. \quad (60)$$

Finally, we conclude that

$$\lim_{t \to \infty} \mathbb{E}\left\{ \|\boldsymbol{x}_t - \boldsymbol{x}^*\|^2 \right\} \leq 2 \lim_{t \to \infty} \mathbb{E}\left\{ \|\boldsymbol{x}_t - \boldsymbol{x}^*(\boldsymbol{\alpha}_{t-1})\|^2 \right\} + 2 \lim_{t \to \infty} \mathbb{E}\left\{ \|\boldsymbol{x}^* - \boldsymbol{x}^*(\boldsymbol{\alpha}_{t-1})\|^2 \right\} \underset{(a)}{=} 0 \quad (61)$$

where (a) uses Lemma 5 for the first term (conditions 1, 3 and 4 imply that $\frac{\varepsilon_t}{\eta_t} \to 0$) and (59) for the second term.