A Appendix

A.1 Proof of Corollary 4.

Proof. Since $(\boldsymbol{x}^{(k)})_{k\in\mathbb{N}}$ converges to $\boldsymbol{x}^*$, there exists $k_0 \geq 0$ such that $\boldsymbol{x}^{(k)}$ lies in $B_{\varepsilon(\boldsymbol{u})}(\boldsymbol{x}^*(\boldsymbol{u}))$ for all $k \geq k_0$. Thus, from the Lipschitz continuity of φ on $\mathcal{Y}$, we have:

$$\left\|\varphi(\boldsymbol{x}^{(k)}, \boldsymbol{u}) - \varphi(\boldsymbol{x}^*, \boldsymbol{u})\right\| \leq C\frac{\kappa + m(\boldsymbol{u})}{m(\boldsymbol{u})^2}\left\|\boldsymbol{x}^{(k)} - \boldsymbol{x}^*\right\|,$$

for all $k \geq k_0$. □

A.2 Proof of Lemma 6.

Proof. Since $f(\cdot, \boldsymbol{u})$ is convex and $\nabla_{\boldsymbol{x}} f$ is $L(\boldsymbol{u})$-Lipschitz continuous on $\mathcal{Z}$, therefore, for all $\boldsymbol{u} \in \mathcal{U}$ and $\alpha \leq 1/L(\boldsymbol{u})$, the first part of the proposition follows from (Bertsekas, 1999) and Induction. In particular we have:

$$f(\boldsymbol{x}^{(k)}, \boldsymbol{u}) - f(\boldsymbol{x}^*(\boldsymbol{u}), \boldsymbol{u}) \leq \frac{1}{2\alpha k}\left\|\boldsymbol{e}^{(0)}\right\|^2 = \mathcal{O}(\frac{1}{k}), \quad (5)$$

for $k \in \mathbb{N}$. Thus the sequence $(\boldsymbol{x}^{(k)})_{k\in\mathbb{N}}$ lies in $\mathcal{X}(\boldsymbol{u})$ and from the continuity of f and Assumption A1, converges to $\boldsymbol{x}^*(\boldsymbol{u})$. This implies that, there exists $\delta(\boldsymbol{u}) > 0$ such that after at most $k_0 \sim \mathcal{O}(1/\delta(\boldsymbol{u}))$ iterations of (GD), the sequence $(\boldsymbol{x}^{(k)})_{k\in\mathbb{N}}$ lies in the set $\mathrm{lev}_{\leq f(\boldsymbol{x}^*,\boldsymbol{u})+\delta(\boldsymbol{u})} f(\cdot, \boldsymbol{u}) \subseteq B_{\varepsilon(\boldsymbol{u})}(\boldsymbol{x}^*)$ and we have for all $k \geq k_0$:

$$\begin{aligned}\boldsymbol{e}^{(k+1)} &= \boldsymbol{e}^{(k)} - \alpha(\nabla_{\boldsymbol{x}} f(\boldsymbol{x}^{(k)}, \boldsymbol{u}) - \nabla_{\boldsymbol{x}} f(\boldsymbol{x}^*(\boldsymbol{u}), \boldsymbol{u})) \\ &= R_g(\boldsymbol{z}^{(k)})\boldsymbol{e}^{(k)}.\end{aligned}$$

Because $\alpha \leq 1/L(\boldsymbol{u})$ and from Equation (1), the term given by:

$$q_{GD}(\boldsymbol{u}) := \sup\{\|R_{GD}(\boldsymbol{x}, \alpha)\| : \boldsymbol{x} \in B_{\varepsilon(\boldsymbol{u})}(\boldsymbol{x}^*(\boldsymbol{u}))\}. \quad (6)$$

lies in $[0, 1)$ and the inequality follows. □

A.3 Proof of Proposition 8.

Proof. We simplify the term $\dot{\boldsymbol{e}}^{(k+1)}$ as:

$$\begin{aligned}\dot{\boldsymbol{e}}^{(k+1)} &= R_{GD}^{(k)}\dot{\boldsymbol{x}}^{(k)} - \alpha\nabla_{\boldsymbol{xu}} f(\boldsymbol{x}^{(k)}, \boldsymbol{u})\dot{\boldsymbol{u}} \\ &\quad - R_{GD}^*\dot{\boldsymbol{x}}^* + \alpha\nabla_{\boldsymbol{xu}} f(\boldsymbol{x}^*, \boldsymbol{u})\dot{\boldsymbol{u}} \\ &= R_{GD}^{(k)}\dot{\boldsymbol{e}}^{(k)} + \big(D(\nabla_{\boldsymbol{x}} f)(\boldsymbol{x}^{(k)}, \boldsymbol{u}) \\ &\quad - D(\nabla_{\boldsymbol{x}} f)(\boldsymbol{x}^*, \boldsymbol{u})\big)(\dot{\boldsymbol{x}}^*, \boldsymbol{s}),\end{aligned}$$

where we assigned $R_{GD}(\boldsymbol{x}^*, \alpha)$ to R_{GD}^*. Rearranging the expression on the right hand side, taking the norm and recursive expansion yields the desired inequality for $k \geq k_0$ and $C_1 := C\|\boldsymbol{s}\|(\kappa + m(\boldsymbol{u}))/m(\boldsymbol{u})$. □

A.4 Proof of Proposition 10.

Proof. The difference of the sequence generated by (GD-FI) with $\varphi(\boldsymbol{x}^{(k)}, \boldsymbol{u})\boldsymbol{s}$ can be simplified as:

$$\hat{\boldsymbol{x}}_K^{(k+1)} - \varphi(\boldsymbol{x}^{(K)}, \boldsymbol{u})\boldsymbol{s} = R_{GD}^{(K)}\big(\hat{\boldsymbol{x}}_K^{(k)} - \varphi(\boldsymbol{x}^{(K)}, \boldsymbol{u})\boldsymbol{s}\big).$$

After taking the norm, expanding the expression on the right recursively and using Equation (6), we arrive at the first inequality. For (GD-RI), we have:

$$\begin{aligned}\tilde{\boldsymbol{u}}_K^{(n+1)} &= \tilde{\boldsymbol{u}}_K^{(n)} - \alpha\tilde{\boldsymbol{x}}^{(K-n)}\nabla_{\boldsymbol{xu}} f \\ &= \tilde{\boldsymbol{u}}_K^{(0)} - \alpha\Big(\sum_{i=0}^{n}\tilde{\boldsymbol{x}}^{(K-n+i)}\Big)\nabla_{\boldsymbol{xu}} f \\ &= -\alpha\tilde{\boldsymbol{x}}^{(K)}\Big(\sum_{i=0}^{n}(R_{HB}^{(K)})^i\Big)\nabla_{\boldsymbol{xu}} f \\ &= -\alpha\boldsymbol{r}^T(I_N - R_{GD}^{(K)})^{-1}\big(I_N - (R_{GD}^{(K)})^{n+1}\big)\nabla_{\boldsymbol{xu}} f \\ &= -\boldsymbol{r}^T\nabla_{\boldsymbol{x}}^2 f^{-1}\big(I_N - (R_{GD}^{(K)})^{n+1}\big)\nabla_{\boldsymbol{xu}} f \\ &= \boldsymbol{r}^T\varphi(\boldsymbol{x}^{(K)}, \boldsymbol{u}) + \boldsymbol{r}^T\nabla_{\boldsymbol{x}}^2 f^{-1}(R_{GD}^{(K)})^{n+1}\nabla_{\boldsymbol{xu}} f.\end{aligned}$$

By taking the norm of the error term $\tilde{\boldsymbol{u}}_K^{(n)} - \boldsymbol{r}^T\varphi(\boldsymbol{x}^{(K)}, \boldsymbol{u})$ from above equation and using Equation (6), we get the second inequality. □

A.5 Proof of Corollary 12.

Proof. $\boldsymbol{x}^{(K)} \in B_{\varepsilon(\boldsymbol{u})}(\boldsymbol{x}^*(\boldsymbol{u}))$ implies $\alpha \leq 1/L(\boldsymbol{u})$ is satisfied for our choice of step size from Equation (1) and (Boyd and Vandenberghe, 2004). Since the conditions of Proposition 10 are satisfied, the proof follows. □

A.6 Proof of Lemma 13.

Proof. For all $\boldsymbol{u} \in \mathcal{U}$ and for given choices of α and β, the first part of the proof follows from (1) and (Polyak, 1987). This implies that $\boldsymbol{x}^{(k)} \in \mathcal{X}(\boldsymbol{u})$ for all $k \in \mathbb{N}$. Also the sequence $(\boldsymbol{x}^{(k)})_{k\in\mathbb{N}}$ converges to $\boldsymbol{x}^*(\boldsymbol{u})$ from the continuity of f and uniqueness of $\boldsymbol{x}^*(\boldsymbol{u})$. Therefore, there exists $k_0 \geq 0$ such that for all $k \geq k_0$ we have $\boldsymbol{x}^{(k)} \in B_{\varepsilon(\boldsymbol{u})}(\boldsymbol{x}^*(\boldsymbol{u}))$. From mean value theorem, the error term $\boldsymbol{e}^{(k+1)}$ is simplified as:

$$\begin{aligned}\boldsymbol{e}^{(k+1)} &= (1+\beta)\boldsymbol{x}^{(k)} - \alpha(\nabla_{\boldsymbol{x}} f(\boldsymbol{x}^{(k)}, \boldsymbol{u}) \\ &\quad - \nabla_{\boldsymbol{x}} f(\boldsymbol{x}^*, \boldsymbol{u})) - \beta\boldsymbol{x}^{(k-1)} - \boldsymbol{x}^* \\ &= R_{HB}(\boldsymbol{z}^{(k)}, \alpha, \beta)\boldsymbol{e}^{(k)} - \beta\boldsymbol{e}^{(k-1)},\end{aligned}$$

for some $\boldsymbol{z}^{(k)} \in \operatorname{conv}\{\boldsymbol{x}^{(k)}, \boldsymbol{x}^*\}$. We assign $\boldsymbol{y}^{(k)} := (\boldsymbol{x}^{(k+1)}, \boldsymbol{x}^{(k)})$ and $\boldsymbol{y}^* := (\boldsymbol{x}^*, \boldsymbol{x}^*)$ and compute the error term for this sequence as:

$$\begin{aligned}\boldsymbol{y}^{(k)} - \boldsymbol{y}^* &= (\boldsymbol{e}^{(k+1)}, \boldsymbol{e}^{(k)}) \\ &= (R_{HB}(\boldsymbol{z}^{(k)}, \alpha, \beta)\boldsymbol{e}^{(k)} - \beta\boldsymbol{e}^{(k-1)}, \boldsymbol{e}^{(k)}) \\ &= T(\boldsymbol{z}^{(k)}, \alpha, \beta)(\boldsymbol{y}^{(k-1)} - \boldsymbol{y}^*)\,,\end{aligned} \tag{7}$$

where we define $T : \mathbb{R}^N \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}^{2N \times 2N}$, a matrix-valued function as:

$$T(\boldsymbol{x}, \alpha, \beta) = \begin{bmatrix} R_{HB}(\boldsymbol{x}, \alpha, \beta) & -\beta I_N \\ I_N & 0_N \end{bmatrix}. \tag{8}$$

Here we use subscripts to denote the order of identitiy and zero matrices to avoid any confusion. Let $\rho(A)$ be the spectral radius of matrix A, then from (Polyak, 1987), (1) and the compactness of our $\varepsilon(\boldsymbol{u})$-neighbourhood, $q_{HB}(\boldsymbol{u})$ defined by:

$$q_{HB}(\boldsymbol{u}) = \sup\{\rho(T(\boldsymbol{x}, \alpha, \beta)) : \boldsymbol{x} \in B_{\varepsilon(\boldsymbol{u})}(\boldsymbol{x}^*(\boldsymbol{u}))\}\,, \tag{9}$$

lies in $[0, 1)$ for every $\boldsymbol{u} \in \mathcal{U}$ and given choices of α and β. From Gelfand's relation between spectral radius and the norm of a matrix (Gelfand, 1941), we arrive at our result by taking the norm of the last identity in (7) and recursively expanding up to k_0. □

A.7 Proof of Proposition 15.

Proof. We assign the expression $R_{HB}(\boldsymbol{x}^*, \alpha, \beta)$ to R^*_{HB} and compute

$$\begin{aligned} R^{(k)}_{HB}\dot{\boldsymbol{x}}^{(k)} - R^*_{HB}\dot{\boldsymbol{x}}^* &= (1+\beta)\dot{\boldsymbol{e}}^{(k)} \\ &\quad - \alpha\big(\nabla^2_{\boldsymbol{x}} f(\boldsymbol{x}^{(k)}, \boldsymbol{u})\dot{\boldsymbol{x}}^{(k)} - \nabla^2_{\boldsymbol{x}} f(\boldsymbol{x}^*, \boldsymbol{u})\dot{\boldsymbol{x}}^*\big) \\ &= R^{(k)}_{HB}\dot{\boldsymbol{e}}^{(k)} - \alpha\big(\nabla^2_{\boldsymbol{x}} f(\boldsymbol{x}^{(k)}, \boldsymbol{u}) \\ &\quad - \nabla^2_{\boldsymbol{x}} f(\boldsymbol{x}^*, \boldsymbol{u})\big)\dot{\boldsymbol{x}}^*\,, \end{aligned}$$

from which we obtain the following error term:

$$\begin{aligned} \dot{\boldsymbol{e}}^{(k+1)} &= R^{(k)}_{HB}\dot{\boldsymbol{x}}^{(k)} - R^*_{HB}\dot{\boldsymbol{x}}^* - \alpha(\nabla_{\boldsymbol{xu}} f(\boldsymbol{x}^{(k)}, \boldsymbol{u}) \\ &\quad - \nabla_{\boldsymbol{xu}} f(\boldsymbol{x}^*, \boldsymbol{u}))\dot{\boldsymbol{u}} - \beta\dot{\boldsymbol{e}}^{(k-1)} \\ &= \begin{bmatrix} R^{(k)}_{HB} & -\beta I_N \end{bmatrix} \dot{\boldsymbol{y}}^{(k-1)} - \alpha\big(D(\nabla_{\boldsymbol{x}} f)(\boldsymbol{x}^{(k)}, \boldsymbol{u}) \\ &\quad - D(\nabla_{\boldsymbol{x}} f)(\boldsymbol{x}^*, \boldsymbol{u})\big)(\dot{\boldsymbol{x}}^*, \dot{\boldsymbol{u}})\,, \end{aligned}$$

where we similarly define $\dot{\boldsymbol{y}}^{(k)} - \dot{\boldsymbol{y}}^* := (\dot{\boldsymbol{e}}^{(k+1)}, \dot{\boldsymbol{e}}^{(k)})$. Thus the error term for this sequence is given by:

$$\begin{aligned} \dot{\boldsymbol{y}}^{(k)} - \dot{\boldsymbol{y}}^* &= T^{(k)}(\dot{\boldsymbol{y}}^{(k-1)} - \dot{\boldsymbol{y}}^*) \\ &\quad - \alpha\big(E^{(k)} - E^*\big)(\dot{\boldsymbol{x}}^*, \dot{\boldsymbol{u}})\,, \end{aligned} \tag{10}$$

where we set $T^{(k)} := T(\boldsymbol{x}^{(k),\alpha,\beta}, \alpha, \beta)$ and define the map $E : \mathbb{R}^N \times \mathbb{R}^P \to \mathcal{L}(\mathbb{R}^N \times \mathbb{R}^N, \mathbb{R}^N \times \mathbb{R}^P)$ as:

$$E(\boldsymbol{x}, \boldsymbol{u}) := \begin{bmatrix} D(\nabla_{\boldsymbol{x}} f)(\boldsymbol{x}, \boldsymbol{u}) \\ 0_{N,N+P} \end{bmatrix}$$

and assign $E(\boldsymbol{x}^{(k)}, \boldsymbol{u})$ to $E^{(k)}$ and $E(\boldsymbol{x}^*, \boldsymbol{u})$ to E^*. Now taking the norm and recursively expanding the term on the right hand side of Equation (10), we arrive at our result by using the same argument we made in the proof of Lemma 13. □

A.8 Proof of Proposition 17.

Proof. We will work through the proof for both sequences in a similar fashion as in Proposition 10. We first consider the forward mode case where the error for $\hat{\boldsymbol{x}}^{(k)}_K$ is given by:

$$\begin{aligned} \hat{\boldsymbol{x}}^{(k+1)}_K - \varphi(\boldsymbol{x}^{(k)}, \boldsymbol{u})\boldsymbol{s} &= R^{(k)}_{HB}\big(\hat{\boldsymbol{x}}^{(n)} - \varphi(\boldsymbol{x}^{(k)}, \boldsymbol{u})\boldsymbol{s}\big) \\ &\quad - \beta\big(\hat{\boldsymbol{x}}^{(n-1)} - \varphi(\boldsymbol{x}^{(k)}, \boldsymbol{u})\boldsymbol{s}\big)\,. \end{aligned}$$

We can use it to compute the error term for $\hat{\boldsymbol{y}}^{(k)}_K := (\hat{\boldsymbol{x}}^{(k+1)}_K, \hat{\boldsymbol{x}}^{(k)}_K)$ as:

$$\begin{aligned} \hat{\boldsymbol{y}}^{(k)}_K - \begin{bmatrix} \varphi(\boldsymbol{x}^{(K)}, \boldsymbol{u})\boldsymbol{s} \\ \varphi(\boldsymbol{x}^{(K)}, \boldsymbol{u})\boldsymbol{s} \end{bmatrix} &= \begin{bmatrix} \hat{\boldsymbol{x}}^{(k+1)}_K - \varphi(\boldsymbol{x}^{(K)}, \boldsymbol{u})\boldsymbol{s} \\ \hat{\boldsymbol{x}}^{(k)}_K - \varphi(\boldsymbol{x}^{(K)}, \boldsymbol{u})\boldsymbol{s} \end{bmatrix} \\ &= \begin{bmatrix} R^{(K)}_{HB} & -\beta I_N \\ I_N & 0_N \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{x}}^{(k)}_K - \varphi(\boldsymbol{x}^{(K)}, \boldsymbol{u})\boldsymbol{s} \\ \hat{\boldsymbol{x}}^{(k-1)} - \varphi(\boldsymbol{x}^{(K)}, \boldsymbol{u})\boldsymbol{s} \end{bmatrix} \\ &= T^{(K)}\left(\hat{\boldsymbol{y}}^{(k-1)}_K - \begin{bmatrix} \varphi(\boldsymbol{x}^{(K)}, \boldsymbol{u})\boldsymbol{s} \\ \varphi(\boldsymbol{x}^{(K)}, \boldsymbol{u})\boldsymbol{s} \end{bmatrix}\right) \\ &= -\big(T^{(K)}\big)^k \begin{bmatrix} \varphi(\boldsymbol{x}^{(K)}, \boldsymbol{u})\boldsymbol{s} \\ \varphi(\boldsymbol{x}^{(K)}, \boldsymbol{u})\boldsymbol{s} \end{bmatrix}, \end{aligned}$$

where in the last equality we used $\hat{\boldsymbol{y}}^{(0)} = (\hat{\boldsymbol{x}}^{(0)}_K, \hat{\boldsymbol{x}}^{(-1)}_K) = 0$. Because $\boldsymbol{x}^{(k)} \in B_{\varepsilon(\boldsymbol{u})}(\boldsymbol{x}^*(\boldsymbol{u}))$, we use the argument provided in the proof of Lemma 13 to arrive at the first inequality.

We now define $\tilde{\boldsymbol{y}}^{(K-n-1)} := (\tilde{\boldsymbol{x}}^{(K-n-1)}, \tilde{\boldsymbol{x}}^{(K-n)})^T$ which is computed for $n = 0, \ldots, K-1$ as:

$$\begin{aligned}\tilde{\boldsymbol{y}}^{(K-n-1)} &= \begin{bmatrix}\tilde{\boldsymbol{x}}^{(K-n-1)}\\ \tilde{\boldsymbol{x}}^{(K-n)}\end{bmatrix}^T \\ &= \begin{bmatrix}\tilde{\boldsymbol{x}}^{(K-n)}R_{HB}^{(K)} - \beta\tilde{\boldsymbol{x}}^{(K-n+1)}\\ \tilde{\boldsymbol{x}}^{(K-n)}\end{bmatrix}^T \\ &= \begin{bmatrix}\tilde{\boldsymbol{x}}^{(K-n)}\\ \tilde{\boldsymbol{x}}^{(K-n+1)}\end{bmatrix}^T \begin{bmatrix}R_{HB}^{(K)} & I_N\\ -\beta I_N & 0_N\end{bmatrix} \\ &= \tilde{\boldsymbol{y}}^{(K-n)}(T^{(K)})^T\,.\end{aligned}$$

We also compute $\tilde{\boldsymbol{v}}_K^{(n+1)} := (\tilde{\boldsymbol{u}}_K^{(n+1)}, \tilde{\boldsymbol{u}}_K^{(n)})^T$ for $n = 0, \ldots, K-1$ as:

$$\begin{aligned}\tilde{\boldsymbol{v}}_K^{(n+1)} &= \begin{bmatrix}\tilde{\boldsymbol{u}}_K^{(n+1)}\\ \tilde{\boldsymbol{u}}_K^{(n)}\end{bmatrix}^T \\ &= \begin{bmatrix}\tilde{\boldsymbol{u}}_K^{(n)} - \alpha\tilde{\boldsymbol{x}}^{(K-n)}\nabla_{\boldsymbol{xu}}f\\ \tilde{\boldsymbol{u}}_K^{(n-1)} - \alpha\tilde{\boldsymbol{x}}^{(K-n+1)}\nabla_{\boldsymbol{xu}}f\end{bmatrix}^T \\ &= \begin{bmatrix}\tilde{\boldsymbol{u}}_K^{(n+1)}\\ \tilde{\boldsymbol{u}}_K^{(n)}\end{bmatrix}^T - \alpha\begin{bmatrix}\tilde{\boldsymbol{x}}^{(K-n-1)}\\ \tilde{\boldsymbol{x}}^{(K-n)}\end{bmatrix}^T\begin{bmatrix}\nabla_{\boldsymbol{xu}}f & 0_{N,P}\\ 0_{N,P} & \nabla_{\boldsymbol{xu}}f\end{bmatrix} \\ &= \tilde{\boldsymbol{v}}_K^{(n)} - \alpha\tilde{\boldsymbol{y}}^{(K-n)}S^{(K)}\,,\end{aligned}$$

where $S : \mathbb{R}^N \times \mathbb{R}^P \to \mathcal{L}(\mathbb{R}^N \times \mathbb{R}^N, \mathbb{R}^P \times \mathbb{R}^P)$ is defined as:

$$S(\boldsymbol{x}, \boldsymbol{u}) = \begin{bmatrix}\nabla_{\boldsymbol{xu}}f(\boldsymbol{x}, \boldsymbol{u}) & 0_{N,P}\\ 0_{N,P} & \nabla_{\boldsymbol{xu}}f(\boldsymbol{x}, \boldsymbol{u})\end{bmatrix}\,,$$

so that $S(\boldsymbol{x}^{(K)}, \boldsymbol{u})$ is assigned to $S^{(K)}$. Putting the expressions for $\tilde{\boldsymbol{v}}_K^{(n+1)}$ and $\tilde{\boldsymbol{y}}^{(K-n-1)}$ together we notice that they are equivalent to those in (GD-RI). We can therefore simplify $\tilde{\boldsymbol{v}}_K^{(n+1)}$ as:

$$\begin{aligned}\tilde{\boldsymbol{v}}_K^{(n+1)} &= \tilde{\boldsymbol{v}}_K^{(n)} - \alpha\tilde{\boldsymbol{y}}^{(K-n)}S^{(K)} \\ &= \tilde{\boldsymbol{v}}_K^{(0)} - \alpha\Big(\sum_{i=0}^{n}\tilde{\boldsymbol{y}}^{(K-n+i)}\Big)S^{(K)} \\ &= -\alpha\tilde{\boldsymbol{y}}^{(K)}\Big(\sum_{i=0}^{n}(T^{(K)^T})^i\Big)S^{(K)} \\ &= -\alpha(\boldsymbol{r}, 0)^T(I_{2N} - T^{(K)^T})^{-1}\big(I_{2N} \\ &\quad - (T^{(K)^T})^{n+1}\big)S^{(K)}\,,\end{aligned}$$

where our starting points are $\tilde{\boldsymbol{v}}_K^{(0)} := 0$ and $\tilde{\boldsymbol{y}}^{(K)} := (\boldsymbol{r}, 0)^T$.

Now in order to compute the inverse of the matrix

$$I_{2N} - T^{(K)^T} = \begin{bmatrix}\alpha\nabla_{\boldsymbol{x}}^2 f - \beta I_N & -I_N\\ \beta I_N & I_N\end{bmatrix}\,,$$

we use the results given in Lu and Shiou (2002)[Theorem 1]. The Schur complement of I_N (bottom right block in the above matrix) is $(\alpha\nabla_{\boldsymbol{x}}^2 f - \beta I_N) - (-I_N)(I_N)^{-1}(\beta)I_N = \alpha\nabla_{\boldsymbol{x}}^2 f$ which is invertible and we have:

$$(\boldsymbol{r}, 0)^T(I_{2N} - T^{(K)^T})^{-1} = \frac{1}{\alpha}(\boldsymbol{r}^T\nabla_{\boldsymbol{x}}^2 f^{-1}, \boldsymbol{r}^T\nabla_{\boldsymbol{x}}^2 f^{-1})^T\,.$$

We can substitute this term in the expression obtained above for $\tilde{\boldsymbol{v}}_K^{(n+1)}$ and obtain

$$\begin{aligned}\tilde{\boldsymbol{v}}_n^{(K)} &= -\begin{bmatrix}\boldsymbol{r}^T\nabla_{\boldsymbol{x}}^2 f^{-1}\\ \boldsymbol{r}^T\nabla_{\boldsymbol{x}}^2 f^{-1}\end{bmatrix}^T\big(I_{2N} - (T^{(K)^T})^n\big)S^{(K)} \\ &= \begin{bmatrix}\boldsymbol{r}^T\varphi(\boldsymbol{x}^{(K)}, \boldsymbol{u})\\ \boldsymbol{r}^T\varphi(\boldsymbol{x}^{(K)}, \boldsymbol{u})\end{bmatrix}^T + \begin{bmatrix}\boldsymbol{r}^T\nabla_{\boldsymbol{x}}^2 f^{-1}\\ \boldsymbol{r}^T\nabla_{\boldsymbol{x}}^2 f^{-1}\end{bmatrix}^T(T^{(K)^T})^n S^{(K)}\,.\end{aligned}$$

Since the matrix $S(\boldsymbol{x}, \boldsymbol{u})$ has same singular values as $\nabla_{\boldsymbol{xu}}f(\boldsymbol{x}, \boldsymbol{u})$, the second inequality follows. □

A.9 Proof of Corollary 19.

Proof. The proof follows from the fact that, in Propositions 10 and 17, we only assume that the estimate $\boldsymbol{x}^{(K)}$ lies in $B_{\varepsilon(\boldsymbol{u})}(\boldsymbol{x}^*(\boldsymbol{u}))$. We don't put any constraint on how it is computed. □