A APPENDIX

Definition 5 (HTO inequality). *Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ such that $\mathbf{y} \in H_s(x)$. Then, according to Definition 3, the following for any $\mathcal{I}_s^{\mathbf{y}}$ is called HT operator inequality:*

$$|x_i| \leq |x_j| \quad \forall \quad i \in (\mathcal{I}_s^{\mathbf{y}})^c \text{ and } j \in \mathcal{I}_s^{\mathbf{y}}. \tag{11}$$

Definition 6 (The IHT algorithm). *The IHT scheme in Algorithm 1 is an iterative process that applies the HT operator on the updated vector found by the gradient descent algorithm.*

Definition 7 (Basic stationary point). *When $\mathbf{x}^*$ is a local minimizer of Problem (2), then $\nabla_{\mathcal{I}_s^*} f(\mathbf{x}^*) = 0$ when $\|\mathbf{x}^*\|_0 = s$ or $\nabla f(\mathbf{x}^*) = 0$ when $\|\mathbf{x}^*\|_0 < s$, Beck & Eldar (2013). Every point satisfying these conditions is called basic stationary point of Problem (2). Points with the aforementioned property are called basic feasible points in Beck & Eldar (2013).*

A.1 PROOF OF CLAIM 1:

Proof. In this part we show why the hard thresholding operator keeps the s largest entries of its input in absolute value sense. First, notice that one can write $\|\mathbf{z} - \mathbf{x}\|_2^2 = \sum_{i=1}^n (z_i - x_i)^2$. Without loss of generality suppose that $\mathbf{x}$ is given such that its entries are in a descending order in terms of absolute value. Then one can write the following:

$$\sum_{i=1}^{n} (z_i - x_i)^2 = \sum_{j=1}^{s} (z_i - x_i)^2 + \sum_{i=s+1}^{n} (z_i - x_i)^2.$$

The optimal solution should have $n - s$ entries whose values are zero. Let z^* be a vector such that the first s entries of z^* be the s largest entries of $\mathbf{x}$ in absolute value sense and the rest be zero. Then one can get the following:

$$\sum_{i=1}^{n} (z_i^* - x_i)^2 = \sum_{i=s+1}^{n} x_i^2.$$

Because $\sum_{i=s+1}^n x_i^2$ is the sum of $n - s$ smallest entries of $\mathbf{x}$, the objective value in (5) would be minimized. Any choice other than a vector $\mathbf{z}^*$ that has the largest s element of $\mathbf{x}$ and zero elsewhere, $\sum_{i=s+1}^n x_i^2$ cannot lead to the minimum of the function value. Hence, $\mathbf{z}^* = H_s(\mathbf{x})$ keeps the s largest entries of $\mathbf{x}$ in absolute value and zero out the rest. □

A.2 PROOF OF THEOREM 1:

Proof. Fix $0 < \gamma \leq \frac{1}{L_s}$, $\mathcal{I}_s^{\mathbf{x}}$ for a given $\mathbf{x} \in C_s$, $\mathbf{y} \in H_s(\mathbf{x} - \gamma \nabla f(\mathbf{x}))$ and $\mathcal{I}_s^{\mathbf{y}}$. Let $\mathcal{I} := \mathcal{I}_s^{\mathbf{x}} \cup \mathcal{I}_s^{\mathbf{y}}$. Clearly, $\mathbf{y}_{\mathcal{I}^c} = \mathbf{x}_{\mathcal{I}^c} = 0$, $\mathbf{x}_{\mathcal{I} \setminus \mathcal{I}_s^{\mathbf{x}}} = 0$, $\mathbf{y}_{\mathcal{I} \setminus \mathcal{I}_s^{\mathbf{x}}} = 0$, and $\mathbf{y}_{\mathcal{I}_s^{\mathbf{y}}} = (\mathbf{x} - \gamma \nabla f(\mathbf{x}))_{\mathcal{I}_s^{\mathbf{y}}} = \mathbf{x}_{\mathcal{I}_s^{\mathbf{y}}} - \gamma \nabla_{\mathcal{I}_s^{\mathbf{y}}} f(\mathbf{x})$. This shows that

$$(\mathbf{y} - \mathbf{x})_{\mathcal{I}^c} = 0, \quad (\mathbf{y} - \mathbf{x})_{\mathcal{I} \setminus \mathcal{I}_s^{\mathbf{y}}} = -\mathbf{x}_{\mathcal{I} \setminus \mathcal{I}_s^{\mathbf{y}}}, \quad (\mathbf{y} - \mathbf{x})_{\mathcal{I}_s^{\mathbf{y}}} = -\gamma \nabla_{\mathcal{I}_s^{\mathbf{y}}} f(\mathbf{x}).$$

Since $f(\mathbf{x})$ is L_s-RSS and both $\|\mathbf{x}\|_0 \leq s$ and $\|\mathbf{y}\|_0 \leq s$, one can write (3). Then, notice that both the inner product and the norm squared in (3) can be partitioned into two terms based as the following:

$$\begin{aligned}
f(\mathbf{y}) &\leq f(x) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle + \frac{L_s}{2} \|\mathbf{y} - \mathbf{x}\|_2^2 \\
&\leq f(x) + \langle \nabla_{\mathcal{I}_s^{\mathbf{y}}} f(\mathbf{x}), (\mathbf{y} - \mathbf{x})_{\mathcal{I}_s^{\mathbf{y}}} \rangle + \frac{L_s}{2} \|(\mathbf{y} - \mathbf{x})_{\mathcal{I}_s^{\mathbf{y}}}\|_2^2 + \langle \nabla_{\mathcal{I} \setminus \mathcal{I}_s^{\mathbf{y}}} f(\mathbf{x}), (\mathbf{y} - \mathbf{x})_{\mathcal{I} \setminus \mathcal{I}_s^{\mathbf{y}}} \rangle + \frac{L_s}{2} \|(\mathbf{y} - \mathbf{x})_{\mathcal{I} \setminus \mathcal{I}_s^{\mathbf{y}}}\|_2^2 \\
&\leq f(x) - \gamma \langle \nabla_{\mathcal{I}_s^{\mathbf{y}}} f(\mathbf{x}), \nabla_{\mathcal{I}_s^{\mathbf{y}}} f(\mathbf{x}) \rangle + \frac{L_s}{2} \| - \gamma \nabla_{\mathcal{I}_s^{\mathbf{y}}} f(\mathbf{x})\|_2^2 - \langle \nabla_{\mathcal{I} \setminus \mathcal{I}_s^{\mathbf{y}}} f(\mathbf{x}), \mathbf{x}_{\mathcal{I} \setminus \mathcal{I}_s^{\mathbf{y}}} \rangle + \frac{L_s}{2} \|\mathbf{x}_{\mathcal{I} \setminus \mathcal{I}_s^{\mathbf{y}}}\|_2^2 \\
&\leq f(x) - \gamma (1 - \frac{\gamma L_s}{2}) \|\nabla_{\mathcal{I}_s^{\mathbf{y}}} f(\mathbf{x})\|_2^2 - \langle \nabla_{\mathcal{I} \setminus \mathcal{I}_s^{\mathbf{y}}} f(\mathbf{x}), \mathbf{x}_{\mathcal{I} \setminus \mathcal{I}_s^{\mathbf{y}}} \rangle + \frac{L_s}{2} \|\mathbf{x}_{\mathcal{I} \setminus \mathcal{I}_s^{\mathbf{y}}}\|_2^2
\end{aligned} \tag{12}$$

Since $0 < \gamma \leq \frac{1}{L_s}$, we have $0 < L_s \leq \frac{1}{\gamma}$. Hence, one can write the following:

$$
\begin{aligned}
&- \langle \nabla_{\mathcal{I}\backslash\mathcal{I}_s^{\mathbf{y}}} f(\mathbf{x}), \mathbf{x}_{\mathcal{I}\backslash\mathcal{I}_s^{\mathbf{y}}} \rangle + \frac{L_s}{2} \|\mathbf{x}_{\mathcal{I}\backslash\mathcal{I}_s^{\mathbf{y}}}\|_2^2 \\
&\leq -\langle \nabla_{\mathcal{I}\backslash\mathcal{I}_s^{\mathbf{y}}} f(\mathbf{x}), \mathbf{x}_{\mathcal{I}\backslash\mathcal{I}_s^{\mathbf{y}}} \rangle + \frac{1}{2\gamma} \|\mathbf{x}_{\mathcal{I}\backslash\mathcal{I}_s^{\mathbf{y}}}\|_2^2 \\
&= \frac{1}{2\gamma} \Big(\|\mathbf{x}_{\mathcal{I}\backslash\mathcal{I}_s^{\mathbf{y}}}\|_2^2 - 2\gamma \langle \nabla_{\mathcal{I}\backslash\mathcal{I}_s^{\mathbf{y}}} f(\mathbf{x}), \mathbf{x}_{\mathcal{I}\backslash\mathcal{I}_s^{\mathbf{y}}} \rangle + \gamma^2 \|\nabla_{\mathcal{I}\backslash\mathcal{I}_s^{\mathbf{y}}} f(\mathbf{x})\|_2^2 \Big) - \frac{\gamma}{2} \|\nabla_{\mathcal{I}\backslash\mathcal{I}_s^{\mathbf{y}}} f(\mathbf{x})\|_2^2 \\
&= \frac{1}{2\gamma} \|\mathbf{x}_{\mathcal{I}\backslash\mathcal{I}_s^{\mathbf{y}}} - \gamma \nabla_{\mathcal{I}\backslash\mathcal{I}_s^{\mathbf{y}}} f(\mathbf{x})\|_2^2 - \frac{\gamma}{2} \|\nabla_{\mathcal{I}\backslash\mathcal{I}_s^{\mathbf{y}}} f(\mathbf{x})\|_2^2.
\end{aligned}
$$

We claim that

$$
\|\mathbf{x}_{\mathcal{I}\backslash\mathcal{I}_s^{\mathbf{y}}} - \gamma \nabla_{\mathcal{I}\backslash\mathcal{I}_s^{\mathbf{y}}} f(\mathbf{x})\|_2^2 \leq \|\gamma \nabla_{\mathcal{I}\backslash\mathcal{I}_s^{\mathbf{x}}} f(\mathbf{x})\|_2^2 = \gamma^2 \|\nabla_{\mathcal{I}\backslash\mathcal{I}_s^{\mathbf{x}}} f(\mathbf{x})\|_2^2. \tag{13}
$$

To show this claim, recall that from the definition of hard thresholding inequality in Definition 5, one can write the following for any $\mathcal{I}_s^{\mathbf{y}}$:

$$
|\mathbf{x}_q - \gamma \nabla_q f(\mathbf{x})| \leq |\mathbf{x}_p - \gamma \nabla_p f(\mathbf{x})|, \quad \forall q \in (\mathcal{I}_s^{\mathbf{y}})^c, \quad \forall p \in \mathcal{I}_s^{\mathbf{y}}
$$

Also, $\mathcal{I}\backslash\mathcal{I}_s^{\mathbf{y}} \subseteq (\mathcal{I}_s^{\mathbf{y}})^c$ and $(\mathcal{I}\backslash\mathcal{I}_s^{\mathbf{x}}) \subseteq \mathcal{I}_s^{\mathbf{y}}$. Therefore, according to the hard thresholding operator for all $i \in \mathcal{I}\backslash\mathcal{I}_s^{\mathbf{y}}$ and for all $j \in \mathcal{I}\backslash\mathcal{I}_s^{\mathbf{x}}$ we have the following:

$$
\begin{aligned}
|\Big(\mathbf{x}_{\mathcal{I}\backslash\mathcal{I}_s^{\mathbf{y}}} - \gamma \nabla_{\mathcal{I}\backslash\mathcal{I}_s^{\mathbf{y}}} f(\mathbf{x})\Big)_i| &\leq |\Big(\mathbf{x}_{\mathcal{I}\backslash\mathcal{I}_s^{\mathbf{x}}} - \gamma \nabla_{\mathcal{I}\backslash\mathcal{I}_s^{\mathbf{x}}} f(\mathbf{x})\Big)_j| \\
&= |\Big(- \gamma \nabla_{\mathcal{I}\backslash\mathcal{I}_s^{\mathbf{x}}} f(\mathbf{x})\Big)_j|
\end{aligned} \tag{14}
$$

where the last inequality follows from the fact that for every entry with index j in $\mathcal{I}\backslash\mathcal{I}_s^{\mathbf{x}}$, the corresponding value is zero. Also, since the number of elements in $|\mathcal{I}_s^{\mathbf{x}}| = |\mathcal{I}_s^{\mathbf{y}}|$, one can write the following:

$$
|\mathcal{I}\backslash\mathcal{I}_s^{\mathbf{y}}| = |\mathcal{I}_s^{\mathbf{x}}| - |\mathcal{I}_s^{\mathbf{x}} \cap \mathcal{I}_s^{\mathbf{y}}| = |\mathcal{I}_s^{\mathbf{y}}| - |\mathcal{I}_s^{\mathbf{x}} \cap \mathcal{I}_s^{\mathbf{y}}| = |\mathcal{I}\backslash\mathcal{I}_s^{\mathbf{x}}| \tag{15}
$$

which implies that the numbers of elements in $\mathcal{I}\backslash\mathcal{I}_s^{\mathbf{y}}$ and $\mathcal{I}\backslash\mathcal{I}_s^{\mathbf{x}}$ are the same. Hence, one can square each inequality in (14) to get the $\|\mathbf{x}_{\mathcal{I}\backslash\mathcal{I}_s^{\mathbf{y}}} - \gamma \nabla_{\mathcal{I}\backslash\mathcal{I}_s^{\mathbf{y}}} f(\mathbf{x})\|_2^2 \leq \|\gamma \nabla_{\mathcal{I}\backslash\mathcal{I}_s^{\mathbf{x}}} f(\mathbf{x})\|_2^2$.

Using (13) one can find an upper bound on the right hand side of (12) as follows:

$$
f(\mathbf{y}) \leq f(\mathbf{x}) - \gamma(1 - \frac{\gamma L_s}{2}) \|\nabla_{\mathcal{I}_s^{\mathbf{y}}} f(\mathbf{x})\|_2^2 + \frac{1}{2\gamma} \gamma^2 \|\gamma \nabla_{\mathcal{I}\backslash\mathcal{I}_s^{\mathbf{x}}} f(\mathbf{x})\|_2^2 - \frac{\gamma}{2} \|\nabla_{\mathcal{I}\backslash\mathcal{I}_s^{\mathbf{y}}} f(\mathbf{x})\|_2^2. \tag{16}
$$

Notice that $\mathcal{I}_s^{\mathbf{y}}$ is the disjoint union of $\mathcal{I}_s^{\mathbf{x}} \cap \mathcal{I}_s^{\mathbf{y}}$ and $\mathcal{I}\backslash\mathcal{I}_s^{\mathbf{x}}$, i.e., $\mathcal{I}_s^{\mathbf{y}} = (\mathcal{I}_s^{\mathbf{x}} \cap \mathcal{I}_s^{\mathbf{y}}) \cup (\mathcal{I}\backslash\mathcal{I}_s^{\mathbf{x}})$. Therefore, $\|\nabla_{\mathcal{I}_s^{\mathbf{y}}} f(\mathbf{x})\|_2^2 = \|\nabla_{\mathcal{I}_s^{\mathbf{x}} \cap \mathcal{I}_s^{\mathbf{y}}} f(\mathbf{x})\|_2^2 + \|\nabla_{\mathcal{I}_s^{\mathbf{x}} \cap \mathcal{I}_s^{\mathbf{y}}} f(\mathbf{x})\|_2^2$. Substituting the right-hand side into prior inequality yields: By adding some positive values to the right hand side of (16) we have:

$$
\begin{aligned}
f(\mathbf{y}) &\leq f(\mathbf{x}) - \gamma(\frac{1}{2} - \frac{\gamma L_s}{2}) \|\nabla_{\mathcal{I}\backslash\mathcal{I}_s^{\mathbf{x}}} f(\mathbf{x})\|_2^2 - \gamma(1 - \frac{\gamma L_s}{2}) \|\nabla_{\mathcal{I}_s^{\mathbf{x}} \cap \mathcal{I}_s^{\mathbf{y}}} f(\mathbf{x})\|_2^2 - \frac{\gamma}{2} \|\nabla_{\mathcal{I}\backslash\mathcal{I}_s^{\mathbf{y}}} f(\mathbf{x})\|_2^2 \\
&\leq f(\mathbf{x}) - \gamma(\frac{1}{2} - \frac{\gamma L_s}{2}) \Big(\|\nabla_{\mathcal{I}\backslash\mathcal{I}_s^{\mathbf{x}}} f(\mathbf{x})\|_2^2 + \|\nabla_{\mathcal{I}_s^{\mathbf{x}} \cap \mathcal{I}_s^{\mathbf{y}}} f(\mathbf{x})\|_2^2 + \|\nabla_{\mathcal{I}\backslash\mathcal{I}_s^{\mathbf{y}}} f(\mathbf{x})\|_2^2 \Big) \\
&= f(\mathbf{x}) - \frac{\gamma}{2}(1 - \gamma L_s) \|\nabla_{\mathcal{I}_s^{\mathbf{x}} \cup \mathcal{I}_s^{\mathbf{y}}} f(\mathbf{x})\|_2^2
\end{aligned}
$$

where the last equations follows from the fact that $\mathcal{I} = \mathcal{I}_s^{\mathbf{x}} \cup \mathcal{I}_s^{\mathbf{y}}$ is the disjoint union of $\mathcal{I}\backslash\mathcal{I}_s^{\mathbf{x}}$, $\mathcal{I}_s^{\mathbf{x}} \cap \mathcal{I}_s^{\mathbf{y}}$, and $\mathcal{I}\backslash\mathcal{I}_s^{\mathbf{y}}$. □

A.3 Proof of Corollary 1:

Proof. Since $\mathbf{y}_{\mathcal{I}^c} = \mathbf{x}_{\mathcal{I}^c} = 0$ and $\mathcal{I}$ is the disjoint union of $\mathcal{I}\backslash\mathcal{I}_s^{\mathbf{y}}$ and $\mathcal{I}_s^{\mathbf{y}}$, one can write the following:

$$
\begin{aligned}
\|\mathbf{y} - \mathbf{x}\|_2^2 &= \|\mathbf{y}_{\mathcal{I}^c} - \mathbf{x}_{\mathcal{I}^c}\|_2^2 + \|\mathbf{y}_{\mathcal{I}\backslash\mathcal{I}_s^{\mathbf{x}}} - \mathbf{x}_{\mathcal{I}\backslash\mathcal{I}_s^{\mathbf{x}}}\|_2^2 + \|\mathbf{y}_{\mathcal{I}\backslash\mathcal{I}_s^{\mathbf{y}}} - \mathbf{x}_{\mathcal{I}\backslash\mathcal{I}_s^{\mathbf{y}}}\|_2^2 + \|\mathbf{y}_{\mathcal{I}_s^{\mathbf{x}} \cap \mathcal{I}_s^{\mathbf{y}}} - \mathbf{x}_{\mathcal{I}_s^{\mathbf{x}} \cap \mathcal{I}_s^{\mathbf{y}}}\|_2^2 \\
&= \|\mathbf{y}_{\mathcal{I}\backslash\mathcal{I}_s^{\mathbf{y}}} - \mathbf{x}_{\mathcal{I}\backslash\mathcal{I}_s^{\mathbf{y}}}\|_2^2 + \|\mathbf{y}_{\mathcal{I}_s^{\mathbf{y}}} - \mathbf{x}_{\mathcal{I}_s^{\mathbf{y}}}\|_2^2 \\
&= \| - \mathbf{x}_{\mathcal{I}\backslash\mathcal{I}_s^{\mathbf{y}}}\|_2^2 + \|\mathbf{y}_{\mathcal{I}_s^{\mathbf{y}}} - \mathbf{x}_{\mathcal{I}_s^{\mathbf{y}}}\|_2^2 \\
&= \|\mathbf{x}_{\mathcal{I}\backslash\mathcal{I}_s^{\mathbf{y}}}\|_2^2 + \|\gamma \nabla_{\mathcal{I}_s^{\mathbf{y}}} f(\mathbf{x})\|_2^2.
\end{aligned} \tag{17}
$$

By applying the reverse triangle inequality on Inequality (13) one can bound $\|\mathbf{x}_{\mathcal{I}\setminus\mathcal{I}_s^{\mathbf{y}}}\|_2^2$ as follows:

$$\|\mathbf{x}_{\mathcal{I}\setminus\mathcal{I}_s^{\mathbf{y}}}\|_2 - \|\gamma\nabla_{\mathcal{I}\setminus\mathcal{I}_s^{\mathbf{y}}} f(\mathbf{x})\|_2 \leq \|\gamma\nabla_{\mathcal{I}\setminus\mathcal{I}_s^{\mathbf{x}}} f(\mathbf{x})\|_2.$$

Hence, one can write the following:

$$\|\mathbf{x}_{\mathcal{I}\setminus\mathcal{I}_s^{\mathbf{y}}}\|_2^2 \leq (\|\gamma\nabla_{\mathcal{I}\setminus\mathcal{I}_s^{\mathbf{y}}} f(\mathbf{x})\|_2 + \|\gamma\nabla_{\mathcal{I}\setminus\mathcal{I}_s^{\mathbf{x}}} f(\mathbf{x})\|_2)^2 \leq 2\|\gamma\nabla_{\mathcal{I}\setminus\mathcal{I}_s^{\mathbf{y}}} f(\mathbf{x})\|_2^2 + 2\|\gamma\nabla_{\mathcal{I}\setminus\mathcal{I}_s^{\mathbf{x}}} f(\mathbf{x})\|_2^2$$

By plugging the above upper bound in place of $\|\mathbf{x}_{\mathcal{I}\setminus\mathcal{I}_s^{\mathbf{y}}}\|_2^2$ in equation 17, we get the following:

$$\begin{aligned}
\|\mathbf{y}-\mathbf{x}\|_2^2 &\leq 2\|\gamma\nabla_{\mathcal{I}\setminus\mathcal{I}_s^{\mathbf{y}}} f(\mathbf{x})\|_2^2 + 2\|\gamma\nabla_{\mathcal{I}\setminus\mathcal{I}_s^{\mathbf{x}}} f(\mathbf{x})\|_2^2 + \|\gamma\nabla_{\mathcal{I}_s^{\mathbf{y}}} f(\mathbf{x})\|_2^2 \\
&= \|\gamma\nabla_{\mathcal{I}\setminus\mathcal{I}_s^{\mathbf{y}}} f(\mathbf{x})\|_2^2 + \|\gamma\nabla_{\mathcal{I}_s^{\mathbf{y}}} f(\mathbf{x})\|_2^2 + \|\gamma\nabla_{\mathcal{I}\setminus\mathcal{I}_s^{\mathbf{y}}} f(\mathbf{x})\|_2^2 + \|\gamma\nabla_{\mathcal{I}\setminus\mathcal{I}_s^{\mathbf{x}}} f(\mathbf{x})\|_2^2 + \|\gamma\nabla_{\mathcal{I}\setminus\mathcal{I}_s^{\mathbf{x}}} f(\mathbf{x})\|_2^2 \\
&= \|\gamma\nabla_{\mathcal{I}} f(\mathbf{x})\|_2^2 + \|\gamma\nabla_{\mathcal{I}\setminus\mathcal{I}_s^{\mathbf{y}}} f(\mathbf{x})\|_2^2 + \|\gamma\nabla_{\mathcal{I}\setminus\mathcal{I}_s^{\mathbf{x}}} f(\mathbf{x})\|_2^2 + \|\gamma\nabla_{\mathcal{I}\setminus\mathcal{I}_s^{\mathbf{x}}} f(\mathbf{x})\|_2^2 \\
&\leq \|\gamma\nabla_{\mathcal{I}} f(\mathbf{x})\|_2^2 + \|\gamma\nabla_{\mathcal{I}} f(\mathbf{x})\|_2^2 + \|\gamma\nabla_{\mathcal{I}} f(\mathbf{x})\|_2^2 \\
&\leq 3\|\gamma\nabla_{\mathcal{I}} f(\mathbf{x})\|_2^2
\end{aligned}$$

Multiplying by $\frac{\gamma}{2}(1 - L_s\gamma)$ one would get the result. □

A.4 Proof of Corollary 2:

Proof. According to (16), one can write the following:

$$\begin{aligned}
\frac{\gamma}{2}(1-\gamma L_s)\|\nabla_{\mathcal{I}\setminus\mathcal{I}_s^{\mathbf{x}}} f(\mathbf{x})\|_2^2 + \gamma(1-\frac{\gamma L_s}{2})\|\nabla_{\mathcal{I}_s^{\mathbf{x}}\cap\mathcal{I}_s^{\mathbf{y}}} f(\mathbf{x})\|_2^2 + \frac{\gamma}{2}\|\nabla_{\mathcal{I}\setminus\mathcal{I}_s^{\mathbf{y}}} f(\mathbf{x})\|_2^2 &\leq f(\mathbf{x}) - f(\mathbf{y}) \\
\gamma(1-\frac{\gamma L_s}{2})\|\nabla_{\mathcal{I}_s^{\mathbf{x}}\cap\mathcal{I}_s^{\mathbf{y}}} f(\mathbf{x})\|_2^2 + \frac{\gamma}{2}\|\nabla_{\mathcal{I}\setminus\mathcal{I}_s^{\mathbf{y}}} f(\mathbf{x})\|_2^2 &\leq f(\mathbf{x}) - f(\mathbf{y}) \\
\frac{\gamma}{2}\|\nabla_{\mathcal{I}_s^{\mathbf{x}}\cap\mathcal{I}_s^{\mathbf{y}}} f(\mathbf{x})\|_2^2 + \frac{\gamma}{2}\|\nabla_{\mathcal{I}\setminus\mathcal{I}_s^{\mathbf{y}}} f(\mathbf{x})\|_2^2 &\leq f(\mathbf{x}) - f(\mathbf{y}) \\
\frac{\gamma}{2}\|\nabla_{\mathcal{I}_s^{\mathbf{x}}} f(\mathbf{x})\|_2^2 &\leq f(\mathbf{x}) - f(\mathbf{y})
\end{aligned} \tag{18}$$

where the second inequality follows from the fact that $\frac{\gamma}{2}(1-\gamma L_s)$ is nonnegative and one can remove $\gamma(1-\frac{\gamma L_s}{2})\|\nabla_{\mathcal{I}\setminus\mathcal{I}_s^{\mathbf{y}}} f(\mathbf{x})\|_2^2$ from the left hand side. The third inequality follows the fact that $\gamma(1-\frac{\gamma L_s}{2}) \geq \frac{\gamma}{2}$ when $0 < \gamma \leq \frac{1}{L_s}$. □

A.5 Proof of Corollary 3:

Proof. From Corollary 2 one can write the following:

$$\frac{\gamma}{2}\|\nabla_{\mathcal{I}_s^k} f(\mathbf{x}^k)\|_2^2 \leq f(\mathbf{x}^k) - f(\mathbf{x}^{k+1}), \quad k \geq 0$$

which means $f(\mathbf{x}^k) \geq f(\mathbf{x}^{k+1})$. Then, $\Big(f(\mathbf{x}^k)\Big)_{k\geq 0}$ is a nonincreasing sequence. Since f is bounded below and $\Big(f(\mathbf{x}^k)\Big)_{k\geq 0}$ is nonincreasing, $\Big(f(\mathbf{x}^k)\Big)_{k\geq 0}$ is a monotone sequence and converges. Now suppose $\mathbf{x}^*$ is an accumulation point of $\big(\mathbf{x}^k\big)_{k\geq 0}$. Thus, there exists a subsequence $\big(\mathbf{x}^{k_j}\big)_{k_j\geq 0}$ of $\big(\mathbf{x}^k\big)_{k\geq 0}$ converging to $\mathbf{x}^*$. Differentiability of f implies its continuity and its continuity implies $\Big(f(\mathbf{x}^{k_j})\Big)_{k_j\geq 0} \to f(\mathbf{x}^*)$. Now, using the fact that $\Big(f(\mathbf{x}^k)\Big)_{k\geq 0}$ is nonincreasing, one can write $f(\mathbf{x}^*) \leq f(\mathbf{x}^{k_j})$ for all $k_j \geq 0$. Thus, $f(\mathbf{x}^*) \leq f(\mathbf{x}^k)$ for all $k \geq 0$ otherwise we get a contradiction. Since, $\big(f(\mathbf{x}^k)\big)_{k\geq 0}$ is monotone and has a convergent subsequence, it converges to the limit point of its subsequence, i.e., $\Big(f(\mathbf{x}^k)\Big)_{k\geq 0} \to f(\mathbf{x}^*)$. □

A.6 PROOF OF THEOREM 2:

Proof. Fix an arbitrary constant $\gamma \in (0, \frac{1}{L_s}]$. Consider $\|\tilde{\mathbf{x}}\|_0 = s$ first. In this case, $\|\tilde{\mathbf{x}}\|_0 = s$. Consequently, $\mathcal{I}_s^{\tilde{\mathbf{x}}}$ is unique and is given by $\text{supp}(\tilde{\mathbf{x}})$. Also, $\min\left(|\tilde{x}_i| : i \in \mathcal{I}_s^{\tilde{\mathbf{x}}}\right) > 0$. Define $\delta_1 := \frac{\min\left(|\tilde{x}_i| : i \in \mathcal{I}_s^{\tilde{\mathbf{x}}}\right)}{2} > 0$. Due to the continuity of $\min$ function and uniqueness of $\mathcal{I}_s^{\tilde{\mathbf{x}}}$ which is equal to $\text{supp}(\tilde{\mathbf{x}})$, there exists a neighborhood $\mathcal{N}_1$ of $\tilde{\mathbf{x}}$ such that $\min\left(|x_i| : i \in \mathcal{I}_s^{\tilde{\mathbf{x}}}\right) > \frac{\min\left(|\tilde{x}_i| : i \in \mathcal{I}_s^{\tilde{\mathbf{x}}}\right)}{2}$ for all $\mathbf{x}$ in $\mathcal{N}_1 := \{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x} - \tilde{\mathbf{x}}\|_2 < \delta_1\}$. Using uniqueness of $\mathcal{I}_s^{\tilde{\mathbf{x}}}$ one can define $h(\mathbf{x})$ as the following:

$$h(\mathbf{x}) := \max\Big(|x_j - \gamma\nabla_j f(\mathbf{x})| : j \notin\in \mathcal{I}_s^{\tilde{\mathbf{x}}}\Big) - \min\Big(|x_i - \gamma\nabla_i f(\mathbf{x})| : i \in \mathcal{I}_s^{\tilde{\mathbf{x}}}\Big).$$

Moreover, $\mathcal{I}_s^{\tilde{\mathbf{x}}} = \text{supp}(\tilde{\mathbf{x}})$ implies $\nabla_{\mathcal{I}_s^{\tilde{\mathbf{x}}}} f(\tilde{\mathbf{x}}) = 0$, $\tilde{\mathbf{x}}_{(\mathcal{I}_s^{\tilde{\mathbf{x}}})^c} = 0$, and $h(\tilde{\mathbf{x}}) > 0$ where

$$h(\tilde{\mathbf{x}}) = \max\Big(|\tilde{x}_j - \gamma\nabla_j f(\tilde{\mathbf{x}})| : j \notin\in \mathcal{I}_s^{\tilde{\mathbf{x}}}\Big) - \min\Big(|\tilde{x}_i - \gamma\nabla_i f(\tilde{\mathbf{x}})| : i \in \mathcal{I}_s^{\tilde{\mathbf{x}}}\Big) > 0.$$

Let $\beta_1 := h(\tilde{\mathbf{x}})$ and $\nu := \frac{3\beta_1^2}{32\gamma} > 0$. Due to the continuity of ∇f, $\min$, and $\max$ functions, $h(\mathbf{x})$ is a continuous function and there exists $\delta_2 > 0$ and a neighborhood $\mathcal{N}_2 = \{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x} - \tilde{\mathbf{x}}\|_2 < \delta_2\}$ such that $h(\mathbf{x}) > \frac{h(\tilde{\mathbf{x}})}{2}$. Let $\mathcal{N} = \{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x} - \tilde{\mathbf{x}}\|_2 < \delta\}$ where $\delta = \min(\delta_1, \delta_2) > 0$. Thus, for all $\mathbf{x} \in \mathcal{N} \cap C_s$, one has $\|\mathbf{x}\|_0 = s$, $\mathcal{I}_s^{\mathbf{x}} = \text{supp}(\mathbf{x}) = \text{supp}(\tilde{\mathbf{x}})$ and

$$\max\Big(|x_j - \gamma\nabla_j f(\mathbf{x})| : j \notin\in \mathcal{I}_s^{\mathbf{x}}\Big) - \min\Big(|x_i - \gamma\nabla_i f(\mathbf{x})| : i \in \mathcal{I}_s^{\mathbf{x}}\Big) \geq \frac{\beta_1}{2}.$$

Fix an arbitrary $\mathbf{x} \in \mathcal{N} \cap C_s$. Then, for any $\mathbf{y} \in H_s(\mathbf{x} - \gamma\nabla f(\mathbf{x}))$ and any $\mathcal{I}_s^{\mathbf{y}}$ there exist two indices $r \in \text{Argmin}\{|x_i - \gamma\nabla_i f(\mathbf{x})| : i \in \mathcal{I}_s^{\mathbf{x}}\}$ and $t \in \text{Argmin}\{|x_j - \gamma\nabla_j f(\mathbf{x})| : j \notin \mathcal{I}_s^{\mathbf{x}}\}$ such that $r \notin \mathcal{I}_s^{\mathbf{y}}$ and $t \in \mathcal{I}_s^{\mathbf{y}}$. Clearly, $r \in \mathcal{I}_s^{\mathbf{x}}$ and $t \notin \mathcal{I}_s^{\mathbf{x}}$. Recall that $\mathcal{I} := \mathcal{I}_s^{\mathbf{x}} \cup \mathcal{I}_s^{\mathbf{y}}$. Thus, $r \in \mathcal{I}\backslash\mathcal{I}_s^{\mathbf{y}}$ and $t \in \mathcal{I}\backslash\mathcal{I}_s^{\mathbf{x}}$. Thus, one has the following:

$$0 \leq |x_r - \gamma\nabla_r f(\mathbf{x})| \leq |x_t - \gamma\nabla_t f(\mathbf{x})| - \frac{\beta_1}{2} = \gamma|\nabla_t f(\mathbf{x})| - \frac{\beta_1}{2} \leq \gamma|\nabla_t f(\mathbf{x})| - \frac{\beta_1}{4}$$

where we used the fact that $x_t = 0$. Observe that the above implies $\gamma|\nabla_t f(\mathbf{x})| \leq \frac{\beta_1}{2}$. Therefore, one can write the following:

$$0 \leq |x_r - \gamma\nabla_r f(\mathbf{x})|^2 \leq (\gamma|\nabla_t f(\mathbf{x})| - \frac{\beta_1}{4})^2 = (\gamma|\nabla_t f(\mathbf{x})|)^2 - \frac{\beta_1}{2}(\gamma|\nabla_t f(\mathbf{x})|) + \frac{\beta_1^2}{16} \leq (\gamma|\nabla_t f(\mathbf{x})|)^2 - \frac{3}{16}\beta_1^2$$

According to (14) for all $i \in \mathcal{I}\backslash\mathcal{I}_s^{\mathbf{y}}$ with $i \neq r$ and for all $j \in \mathcal{I}\backslash\mathcal{I}_s^{\mathbf{x}}$ with $j \neq t$, one has $|x_i - \gamma\nabla_i f(\mathbf{x})| \leq |\nabla_j f(\mathbf{x})|$. Also, according to (15) one can write the following:

$$\|\mathbf{x}_{\mathcal{I}\backslash\mathcal{I}_s^{\mathbf{y}}} - \gamma\nabla_{\mathcal{I}\backslash\mathcal{I}_s^{\mathbf{y}}} f(\mathbf{x})\|_2^2 \leq \|\gamma\nabla_{\mathcal{I}\backslash\mathcal{I}_s^{\mathbf{x}}} f(\mathbf{x})\|_2^2 - \frac{3}{16}\beta_1^2.$$

Therefore, using (16) and applying the same step as the we did for the last step of proof of Theorem 1 one can write the following:

$$\begin{aligned} f(\mathbf{y}) &\leq f(\mathbf{x}) - \gamma(1 - \frac{\gamma L_s}{2})\|\nabla_{\mathcal{I}_s^{\mathbf{y}}} f(\mathbf{x})\|_2^2 + \frac{1}{2\gamma}(\|\gamma\nabla_{\mathcal{I}\backslash\mathcal{I}_s^{\mathbf{x}}} f(\mathbf{x})\|_2^2 - \frac{3}{16}\beta_1^2) - \frac{\gamma}{2}\|\nabla_{\mathcal{I}\backslash\mathcal{I}_s^{\mathbf{y}}} f(\mathbf{x})\|_2^2 \\ &\leq f(\mathbf{x}) - \frac{\gamma}{2}(1 - \gamma L_s)\|\nabla_{\mathcal{I}_s^{\mathbf{x}} \cup \mathcal{I}_s^{\mathbf{y}}} f(\mathbf{x})\|_2^2 - \frac{3}{32}\beta_1^2 \\ &\leq f(\mathbf{x}) - \nu \end{aligned}$$

where we used $\nu := \frac{3\beta_1^2}{32\gamma} > 0$.

Consider $\|\tilde{\mathbf{x}}\|_0 < s$ next. In this case, $\text{supp}(\tilde{\mathbf{x}})$ is a proper subset of any $\mathcal{I}_s^{\tilde{\mathbf{x}}}$, i.e., $\text{supp}(\tilde{\mathbf{x}}) \subset \mathcal{I}_s^{\tilde{\mathbf{x}}}$. Hence,

$$\min\Big(|\tilde{x}_i| : i \in \mathcal{I}_s^{\tilde{\mathbf{x}}}\Big) = 0, \quad \forall \mathcal{I}^{\tilde{\mathbf{x}}_s}; \text{ and } \max\Big(|\tilde{x}_j - \gamma\nabla_j f(\tilde{\mathbf{x}})| : j \notin \text{supp}(\tilde{\mathbf{x}})\Big) > 0. \tag{19}$$

Let $\|\tilde{\mathbf{x}}\|_0 = d < s$. Since $\|\tilde{\mathbf{x}}\|_0 = d < s$, there are $\binom{n-d}{s-d}$ sets of $\mathcal{I}_s^{\tilde{\mathbf{x}}}$. The elements of gradients over these sets of $\mathcal{I}_s^{\tilde{\mathbf{x}}}$ are either zero or nonzero. Thus, define the following two (finite) families of $\mathcal{I}_s^{\tilde{\mathbf{x}}}$'s which constitute a disjoint union of all $\mathcal{I}_s^{\tilde{\mathbf{x}}}$:

$$\mathbb{I}_1 = \{\mathcal{I}_s^{\tilde{\mathbf{x}}} \mid \nabla_{\mathcal{I}_s^{\tilde{\mathbf{x}}}} f(\tilde{\mathbf{x}}) \neq 0\}, \quad \text{and} \quad \mathbb{I}_2 = \{\mathcal{I}_s^{\tilde{\mathbf{x}}} \mid \nabla_{\mathcal{I}_s^{\tilde{\mathbf{x}}}} f(\tilde{\mathbf{x}}) = 0\}.$$

Clearly, $\mathbb{I}_1$ is nonempty because (19) implies that there exists an $\mathcal{I}_s^{\tilde{\mathbf{x}}}$ that contains $j \in \text{Argmin}\{|x_j - \gamma \nabla_j f(\mathbf{x})| : j \notin \text{supp}(\tilde{\mathbf{x}})\}$ such that $|\gamma \nabla_j f(\mathbf{x})| \neq 0$. If $\mathbb{I}_2$ is empty, define

$$\beta_{2,1} := \min\Big(\|\gamma \nabla_{\mathcal{I}_s^{\tilde{\mathbf{x}}}} f(\tilde{\mathbf{x}})\| \mid \mathcal{I}_s^{\tilde{\mathbf{x}}} \in \mathbb{I}_1\Big)$$

and let $\beta_2 := \beta_{2,1} > 0$. If $\mathbb{I}_2$ is nonempty, then for any $\mathcal{I}_s^{\tilde{\mathbf{x}}} \in \mathbb{I}_2$, we have,

$$0 = \min\Big(|\tilde{x}_i - \gamma \nabla_i f(\tilde{\mathbf{x}})| : i \in \mathcal{I}_s^{\tilde{\mathbf{x}}}\Big) < \max\Big(|\tilde{x}_j - \gamma \nabla_j f(\tilde{\mathbf{x}})| : j \notin \mathcal{I}_s^{\tilde{\mathbf{x}}}\Big)$$

because $\text{supp}(\tilde{\mathbf{x}}) \subset \mathcal{I}_s^{\tilde{\mathbf{x}}}$. If $\mathbb{I}_2$ is empty, we let $\beta_2 := \beta_{2,1} > 0$; otherwise, define

$$\beta_{2,2} := \min\Bigg(\max\Big(|\tilde{x}_j - \gamma \nabla_j f(\tilde{\mathbf{x}})| : j \notin \mathcal{I}_s^{\tilde{\mathbf{x}}}\Big) - \min\Big(|\tilde{x}_i - \gamma \nabla_i f(\tilde{\mathbf{x}})| : i \in \mathcal{I}_s^{\tilde{\mathbf{x}}}\Big) \mid \mathcal{I}_s^{\tilde{\mathbf{x}}} \in \mathbb{I}_2\Bigg) > 0$$

and $\beta := \min(\beta_{2,1}, \beta_{2,2}) > 0$.

Based on emptiness or non-emptiness of $\mathbb{I}_2$ we consider two sub-cases as follows:

(i) Suppose $\mathbb{I}_2$ is empty. For this case define $\nu := \frac{\beta_2^2}{8\gamma} > 0$. Similar to case $\|\tilde{\mathbf{x}}\|_0 = s$, by observing the fact that $\min\big(|\tilde{x}_i| : i \in \text{supp}(\tilde{\mathbf{x}})\big) > 0$ and using the continuity of $\nabla f(\cdot)$ and considering that $\beta_2 = \min\Big(\|\gamma \nabla_{\mathcal{I}_s^{\tilde{\mathbf{x}}}} f(\tilde{\mathbf{x}})\| \mid \mathcal{I}_s^{\tilde{\mathbf{x}}} \in \mathbb{I}_1\Big)$, one can can show that there exists a neighborhood $\mathcal{N}$ of $\tilde{\mathbf{x}}$ such that for all $\mathbf{x} \in \mathcal{N} \cap C_s$, $\|\nabla_{\mathcal{I}_s^{\tilde{\mathbf{x}}}} f(\mathbf{x})\| \geq \frac{\beta_2}{2\gamma}$ for any $\mathcal{I}_s^{\tilde{\mathbf{x}}} \in \mathbb{I}_1$. Also, for all $\mathbf{x} \in \mathcal{N} \cap C_s$, $\text{supp}(\mathbf{x}) = \text{supp}(\tilde{\mathbf{x}})$, and $\mathcal{I}_s^{\mathbf{x}} = \mathcal{I}_s^{\tilde{\mathbf{x}}}$ for some $\mathcal{I}_s^{\tilde{\mathbf{x}}}$. Therefore, for all $\mathbf{x} \in \mathcal{N} \cap C_s$ and any $\mathcal{I}_s^{\tilde{\mathbf{x}}}, \mathcal{I}_s^{\tilde{\mathbf{x}}} \in \mathbb{I}_1$ and $\|\nabla_{\mathcal{I}_s^{\mathbf{x}}} f(\mathbf{x})\| \geq \frac{\beta_2}{2\gamma}$. Hence, following Corollary 2 one can write the following for all $\mathbf{x} \in \mathcal{N} \cap C_s$ and any $\mathbf{y} \in H_s(\mathbf{x} - \gamma \nabla f(\mathbf{x}))$:

$$f(\mathbf{x}) - f(\mathbf{y}) \geq \frac{\gamma}{2} \|\nabla_{\mathcal{I}_s^x} f(\mathbf{x})\|_2^2 \geq \frac{\beta_2^2}{8\gamma} = \nu$$

(ii) Suppose $\mathbb{I}_2$ is nonempty. Let $\nu := \frac{3\beta_1^2}{32\gamma} > 0$. In this case there exists a neighborhood $\mathcal{N}$ of $\tilde{\mathbf{x}}$ such that for all $\mathbf{x} \in \mathcal{N} \cap C_s$, $\text{supp}(\mathbf{x}) = \text{supp}(\tilde{\mathbf{x}})$, each $\mathcal{I}_s^{\mathbf{x}}$ equals to some $\mathcal{I}_s^{\tilde{\mathbf{x}}}$, $\|\nabla_{\mathcal{I}_s^{\tilde{\mathbf{x}}}} f(\mathbf{x})\| \geq \frac{\beta_{2,1}}{2\gamma}$ for every $\mathcal{I}_s^{\tilde{\mathbf{x}}} \in \mathbb{I}_1$, and for all $\mathcal{I}_s^{\tilde{\mathbf{x}}} \in \mathbb{I}_2$ one can write the following:

$$\max\Big(|x_j - \gamma \nabla_j f(\mathbf{x})| : j \notin \mathcal{I}_s^{\mathbf{x}})\Big) - \min\Big(|x_i - \gamma \nabla_i f(\mathbf{x})| : i \in \mathcal{I}_s^{\tilde{\mathbf{x}}}\Big) \geq \frac{\beta_{2,2}}{2} > 0.$$

Hence, for all $\mathbf{x} \in \mathcal{N} \cap C_s$ and any $\mathcal{I}_s^{\mathbf{x}}$, either $\mathcal{I}_s^{\mathbf{x}} \in \mathbb{I}_1$ or $\mathcal{I}_s^{\mathbf{x}} \in \mathbb{I}_2$. For the former, we see via the same argument for sub-case (i) that for every $\mathbf{y} \in H_s(\mathbf{x} - \gamma \nabla f(\mathbf{x}))$, $f(\mathbf{x}) - f(\mathbf{y}) \geq \frac{\gamma}{2} \|\nabla_{\mathcal{I}_s^x} f(\mathbf{x})\|_2^2 \geq \frac{\beta_{2,1}^2}{8\gamma} \geq \frac{\beta_2^2}{8\gamma} > \frac{3\beta_2^2}{32\gamma} = \nu$. For the latter, it follows from the similar argument for the case of $\|\tilde{\mathbf{x}}\|_0 = s$ that for any $\mathbf{y} \in H_s(\mathbf{x} - \gamma \nabla f(\mathbf{x}))$, $f(\mathbf{y}) \leq f(\mathbf{x}) - \frac{3\beta_2^2}{32\gamma} = \nu$. This leads to the desired results. □

A.7 Proof of Theorem 3:

Proof. The proof of a) is given in (Beck & Eldar, 2013, Theorem 2.2). To show b) first one needs to show $\mathbf{x}^*$ is a *HT-stable* stationary point. To show that suppose it is not. Then, it is a *HT-unstable* stationary point. According to Theorem 2, there exist $\nu > 0$ and a neighborhood $\mathcal{N}$ of $\mathbf{x}^*$ such that $f(\mathbf{y}) \leq f(\mathbf{x}) - \nu$ for all $\mathbf{x} \in \mathcal{N} \cap C_s$ and any $\mathbf{y} \in P_{C_s}(\mathbf{x} - \gamma \nabla f(\mathbf{x}))$. Let $\mathbf{x} = \mathbf{x}^*$ to get $f(\mathbf{y}) \leq f(\mathbf{x}^*) - \nu$. This contradicts our assumption that $\mathbf{x}^*$ is a global minimizer. Hence, $\mathbf{x}^*$ is a

HT-stable stationary point. Consequently, $\nabla_{\text{supp}(\mathbf{x}^*)} f(\mathbf{x}^*) = 0$. Now, let $\gamma = \frac{1}{L_s}$ and suppose that $\|\mathbf{x}^*\|_0 < s$. Since $\mathbf{x}^*$ is a *HT-stable* point, one can write the following:

$$\min\left(|x_i^*| : i \in \mathcal{I}_s^{\mathbf{x}^*}\right) \geq \gamma \max\left(|\nabla_j f(\mathbf{x}^*)| : j \notin \text{supp}(\mathbf{x}^*)\right).$$

Because $\text{supp}(\mathbf{x}^*)$ is a proper subset of $\mathcal{I}_s^{\mathbf{x}^*}$, $\min\left(|x_i^*| : i \in \mathcal{I}_s^{\mathbf{x}^*}\right) = 0$ which implies $\nabla_{(\text{supp}(\mathbf{x}^*))^c} f(\mathbf{x}^*) = 0$. Together with $\nabla_{\text{supp}(\mathbf{x}^*)} f(\mathbf{x}^*) = 0$, one can conclude $\nabla f(\mathbf{x}^*) = 0$. Thus, $\mathbf{x}^* = \mathbf{x}^* - \gamma \nabla f(\mathbf{x}^*) = P_s(\mathbf{x}^* - \gamma \nabla f(\mathbf{x}^*))$. Now suppose that $\|\mathbf{x}^*\|_0 = s$. Thus, $\text{supp}(\mathbf{x}^*) = \mathcal{I}_s^{\mathbf{x}^*}$. Since $\mathbf{x}^*$ is a *HT-stable* stationary point $\nabla_{\mathcal{I}_s^{\mathbf{x}^*}} f(\mathbf{x}^*) = 0$ and one can write the following:

$$\min\left(|x_i^* - \gamma \nabla_i f(\mathbf{x}^*)| : i \in \mathcal{I}_s^{\mathbf{x}^*}\right) \geq \max\left(|x_j^* - \gamma \nabla_j f(\mathbf{x}^*)| : j \notin \mathcal{I}_s^{\mathbf{x}^*}\right)$$

which implies the following that is definition of HT operator in Definition 5:

$$|x_i^* - \gamma \nabla_i f(\mathbf{x}^*)| \geq |x_j^* - \gamma \nabla_j f(\mathbf{x}^*)| \quad \forall i \in \mathcal{I}_s^{\mathbf{x}^*}, j \notin \mathcal{I}_s^{\mathbf{x}^*}.$$

Thus, $\mathbf{x}^* \in P_s(\mathbf{x}^* - \gamma \nabla f(\mathbf{x}^*))$. □

A.8 Proof of Theorem 4:

Proof. Let $\gamma \in (0, \frac{1}{L_s}]$. Since $\mathbf{x}^*$ is an accumulation point, there exists a subsequence $(\mathbf{x}^{k_j})$ of $(\mathbf{x}^k)$ converging to $\mathbf{x}^*$. Also, there exists a subsequence $(\mathbf{x}^{k_{j_l}})$ of $(\mathbf{x}^{k_j})$ converging to $\mathbf{x}^*$ such that $\mathcal{I}_s^{\mathbf{x}^{k_{j_l}}}$ is a constant set for all k. Let $\mathcal{J} = \mathcal{I}_s^{\mathbf{x}^{k_{j_l}}}$. According to Corollary 2 one can write the following:

$$\frac{\gamma}{2} \|\nabla_{\mathcal{J}} f(\mathbf{x}^{k_{j_l}})\|_2^2 \leq f(\mathbf{x}^{k_{j_l}}) - f(\mathbf{x}^{k_{j_l}+1}).$$

We can sum over l to get the following:

$$\frac{\gamma}{2} \sum_{l=1}^{\infty} \|\nabla_{\mathcal{J}} f(\mathbf{x}^{k_{j_l}})\|_2^2 \leq \sum_{l=1}^{\infty} \left(f(\mathbf{x}^{k_{j_l}}) - f(\mathbf{x}^{k_{j_l}+1})\right) \leq \sum_{k=0}^{\infty} \left(f(\mathbf{x}^k) - f(\mathbf{x}^{k+1})\right)$$

where the second inequality follows the fact that $(\mathbf{x}^{k_{j_l}})$ is a subsequence of $(\mathbf{x}^k)$ and $(f(\mathbf{x}^k))$ is a nonincreasing sequence. Thus, $\frac{\gamma}{2} \sum_{l=1}^{\infty} \|\nabla_{\mathcal{J}} f(\mathbf{x}^{k_{j_l}})\|_2^2 \leq f(\mathbf{x}^0) - f(\mathbf{x}^{k+1})$. By letting l go to infinity, the right hand side would be bounded since f is bounded below on C_s. Therefore, $\nabla_{\mathcal{J}} f(\mathbf{x}^*) = 0$. Notice that $\mathbf{x}^{k_{j_l}}_{(\mathcal{I}_s^{\mathbf{x}^{k_{j_l}}})^c} = 0$ for all k_{j_l}. Since $\mathcal{J} = \mathcal{I}_s^{\mathbf{x}^{k_{j_l}}}$, one has $\mathbf{x}^{k_{j_l}}_{\mathcal{J}^c} = 0$ for all k_{j_l}. On the other hand, $\mathbf{x}^{k_{j_l}} \to \mathbf{x}^*$ so $\mathbf{x}^*_{\mathcal{J}^c} = 0$. This shows $\text{supp}(\mathbf{x}^*) \subseteq \mathcal{J}$. Since $\text{supp}(\mathbf{x}^*) \subseteq \mathcal{J}$, one has $\nabla_{\text{supp}(\mathbf{x}^*)} f(\mathbf{x}^*) = 0$. Now suppose to the contrary that $\mathbf{x}^*$ is a *HT-unstable*. Then, according to Theorem 2, there exists a constant $\nu > 0$ and a neighborhood $\mathcal{N}$ of $\mathbf{x}^*$ such that $f(\mathbf{y}) \leq f(\mathbf{x}) - \nu$ for all $\mathbf{x} \in \mathcal{N} \cap C_s$ and any $\mathbf{y} \in H_s(\mathbf{x} - \gamma \nabla f(\mathbf{x}))$. Thus, there exists $k \geq N$ such that $(\mathbf{x}^{k_{j_l}}) \in \mathcal{N} \cap C_s$ and one can write $f(\mathbf{x}^{k_{j_l}+1}) \leq f(\mathbf{x}^{k_{j_l}}) - \nu$. Then one can sum over j's to get $f(\mathbf{x}^{k_j+1}) - f(\mathbf{x}^{k_1}) \leq -\nu j$. By letting j go to infinity, we get $f(\mathbf{x}^{k_j+1}) \to \infty$ which implies $f(\mathbf{x}^k) \to \infty$. This contradicts the boundedness of f from below. Hence, $\mathbf{x}^*$ is a *HT-stable* stationary point. □

A.9 Proof of Corollary 4:

Proof. Let $\mathbf{x}^*$ be a *HT-unstable* point associated with some $\gamma \in (0, \frac{1}{L_s}]$. Then, according to Theorem 2, there exists a constant $\nu > 0$ and a neighborhood $\mathcal{N}$ of $\mathbf{x}^*$ such that $f(\mathbf{y}) \leq f(\mathbf{x}) - \nu$ for all $\mathbf{x} \in \mathcal{N} \cap C_s$ and any $\mathbf{y} \in H_s(\mathbf{x} - \gamma \nabla f(\mathbf{x}))$. Let $\mathbf{x} = \mathbf{x}^*$ and $\alpha = f(\mathbf{x}^*)$. Thus, $S = \{x \in C_s | f(\mathbf{x}) \leq f(\mathbf{x}^*) = \alpha\}$ is nonempty since $f(\mathbf{y}) \leq f(\mathbf{x}^*) - \nu < f(\mathbf{x}^*)$. Also, let $\left(\mathbf{x}^k\right)_{k \geq 0}$ be an IHT sequence with $\mathbf{x}^0 = \mathbf{y}$. Since $\left(f(\mathbf{x}^k)\right)_{k \geq 0}$ is nonincreasing, $\left(\mathbf{x}^k\right)_{k \geq 0}$ is in S. Because S is bounded, there exists a subsequence $(\mathbf{x}^{k_j})$ of $(\mathbf{x}^k)$ converging to $\tilde{\mathbf{x}}^*$. Since $\tilde{\mathbf{x}}^*$ is an accumulation point of the IHT sequence, by virtue of Theorem 4 it is a *HT-stable* stationary point. Hence, by the continuity of f, one has $f(\mathbf{x}^{k_j}) \to f(\tilde{\mathbf{x}}^*) \leq \mathbf{y} < f(\mathbf{x}^*)$ which is the desired result. □

A.10 PROOF OF COROLLARY 5:

Proof. Fix an arbitrary $0 < \gamma(0, \frac{1}{L_s}]$. By virtue of the proof for Corollary 4 any IHT sequence $\left(\mathbf{x}^k\right)_{k\geq 0}$ is bounded and attains an accumulation point. Suppose *A.1* holds. To show the convergence of $\left(\mathbf{x}^k\right)_{k\geq 0}$, we show that $\left(\mathbf{x}^k\right)_{k\geq 0}$ has a unique accumulation point. Note that due to Theorem 4 any accumulation point of IHT sequence is *HT-stable* stationary point. Now, suppose the accumulation is not unique. Then, $\left(\mathbf{x}^k\right)_{k\geq 0}$ has (at least) two distinct accumulation points denoted by $\mathbf{x}^*$ and $\mathbf{y}^*$, respectively. And, there exist two subsequences $\left(\mathbf{x}^{k_j}\right)$ and $\left(\mathbf{x}^{k_l}\right)$ of $\left(\mathbf{x}^k\right)$ converging to $\mathbf{x}^*$ and $\mathbf{y}^*$, respectively. Since f is continuous, $\left(f(\mathbf{x}^{k_j})\right)$ and $\left(f(\mathbf{x}^{k_l})\right)$ converge to $f(\mathbf{x}^*)$ and $f(\mathbf{y}^*)$, respectively. However, by invoking Corollary 3, one can observe that the sequence of the objective function value $\left(f(\mathbf{x}^k)\right)_{k\geq 0}$ converges. Thus $\left(f(\mathbf{x}^k)\right)_{k\geq 0}$ converges to both $f(\mathbf{x}^*)$ and $f(\mathbf{y}^*)$. This implies that $f(\mathbf{x}^*) = f(\mathbf{y}^*)$, a contradiction. Hence, $\left(\mathbf{x}^k\right)_{k\geq 0}$ has exactly one accumulation point and is convergent. The convergence results under *A.2* follows from $\|\mathbf{x}^{k+1} - \mathbf{x}^k\|_2 \to 0$ as $k \to \infty$ when $0 < \gamma < \frac{1}{L_s}$ (Moré & Sorensen, 1983, Lemma 4.10). □

A.11 PROOF OF COROLLARY 6:

Proof. According to Theorem 2, there exists a constant $\nu > 0$ and a neighborhood $\mathcal{N}$ of $\mathbf{x}^*$ such that $f(\mathbf{y}) \leq f(\mathbf{x}) - \nu$ for all $\mathbf{x} \in \mathcal{N} \cap C_s$ and any $\mathbf{y} \in H_s(\mathbf{x} - \gamma\nabla f(\mathbf{x}))$. Then every IHT sequence with an arbitrary $\mathbf{x}^0 \in C_s$ has finitely many points in $\mathcal{N} \cap C_s$. Otherwise there exists $\mathbf{x}^0 \in C_s$ and an IHT sequence starting from $\mathbf{x}^0$ such that for all $N \in \mathbb{N}$, there exists $k \geq N$ for which $\mathbf{x}^k \in \mathcal{N} \cap C_s$. Then, there exists a subsequence $(\mathbf{x}^{k_j})$ of $(\mathbf{x}^k)$ that is in $\mathcal{N} \cap C_s$. Thus, $f(\mathbf{x}^{k_j+1}) \leq f(\mathbf{x}^{k_j}) - \nu$ for all $j \geq 1$. Then one can sum over j's to get $f(\mathbf{x}^{k_j+1}) - f(\mathbf{x}^{k_1}) \leq -\nu j$. By letting j go to infinity, we get $f(\mathbf{x}^{k_j+1}) \to \infty$ which implies $f(\mathbf{x}^k) \to \infty$. This contradicts the boundedness of f from below. Hence, the claim. □

A.12 PROOF OF PROPOSITION 1:

Proof. Fix an arbitrary constant $\gamma \in (0, \frac{1}{L_s}]$. When $\mathbf{x}^*$ is strictly *HT-stable* point, $\|\mathbf{x}^*\|_0 = s$. Thus, $\min\left(|x_i^*| : i \in \mathcal{I}_s^{\mathbf{x}^*}\right) > 0$. Since $\|\mathbf{x}^*\|_0 = s$, $\mathcal{I}_s^{\mathbf{x}^*}$ is unique and is given by $\text{supp}(\mathbf{x}^*)$. Define $\delta_1 := \frac{\min\left(|x_i^*| : i \in \mathcal{I}_s^{\mathbf{x}^*}\right)}{2} > 0$. Due to the continuity of $\min$ function and uniqueness of $\mathcal{I}_s^{\mathbf{x}^*}$, we know that there exists a neighborhood $\mathcal{N}_1$ of $\mathbf{x}$ such that $\min\left(|x_i| : i \in \mathcal{I}_s^{\mathbf{x}}\right) > \frac{\min\left(|x_i^*| : i \in \mathcal{I}_s^{\mathbf{x}^*}\right)}{2}$ for all $\mathbf{x}$ in $\mathcal{N}_1 := \{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x} - \mathbf{x}^*\|_2 < \delta_1\}$. Also, using uniqueness of $\mathcal{I}_s^{\mathbf{x}^*}$ one can define $h(\mathbf{x})$ as the following:

$$h(\mathbf{x}) := \min\left(|x_i - \gamma\nabla_i f(\mathbf{x})| : i \in \mathcal{I}_s^{\mathbf{x}^*}\right) - \max\left(|x_j - \gamma\nabla_j f(\mathbf{x})| : j \notin \mathcal{I}_s^{\mathbf{x}^*}\right).$$

Moreover, $\mathcal{I}_s^{\mathbf{x}^*} = \text{supp}(\mathbf{x}^*)$ implies $\nabla_{\mathcal{I}_s^{\mathbf{x}^*}} f(\mathbf{x}^*) = 0$, $\mathbf{x}^*_{(\mathcal{I}_s^{\mathbf{x}^*})^c} = 0$, so $h(\mathbf{x}^*) > 0$ where

$$h(\mathbf{x}^*) = \min\left(|x_i - \gamma\nabla_i f(\mathbf{x}^*)| : i \in \mathcal{I}_s^{\mathbf{x}*}\right) - \max\left(|x_j^* - \gamma\nabla_j f(\mathbf{x})| : j \notin \mathcal{I}_s^{\mathbf{x}^*}\right) > 0.$$

Due to continuity of ∇f, $\min$, and $\max$ functions, $h(\mathbf{x})$ is a continuous function and there exists $\delta_2 > 0$ and a neighborhood $\mathcal{N}_2 := \{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x} - \mathbf{x}^*\|_2 < \delta_2\}$ such that $h(\mathbf{x}) > \frac{h(\mathbf{x}^*)}{2}$. Since $\min\left(|x_i^*| : i \in \mathcal{I}_s^{\mathbf{x}^*}\right) > 0$, $\nabla_{\mathcal{I}_s^{\mathbf{x}^*}} f(\mathbf{x}^*) = 0$, $\mathcal{I}_s^{\mathbf{x}^*}$ is unique, and ∇f is continuous there exists $\delta_3 > 0$ and a neighborhood $\mathcal{N}_3 := \{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x} - \mathbf{x}^*\|_2 < \delta_3\}$ such that $\gamma\|\nabla_{\mathcal{I}_s^{\mathbf{x}^*}} f(\mathbf{x}^*)\|_2 < \frac{\min\left(|x_i^*| : i \in \mathcal{I}_s^{\mathbf{x}^*}\right)}{4}$. Let $\mathcal{B} = \{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x} - \mathbf{x}^*\|_2 < \delta\}$ where $\delta = \min(\delta_1, \delta_2, \delta_3) > 0$. Thus, for all $\mathbf{x} \in \mathcal{B} \cap C_s$, one has $\|\mathbf{x}\|_0 = s$, $\mathcal{I}_s^{\mathbf{x}} = \text{supp}(\mathbf{x}) = \text{supp}(\mathbf{x}^*)$ and

$$\min\left(|x_i - \gamma\nabla_i f(\mathbf{x})| : i \in \mathcal{I}_s^{\mathbf{x}}\right) > \max\left(|x_j - \gamma\nabla_j f(\mathbf{x})| : j \notin \mathcal{I}_s^{\mathbf{x}}\right). \tag{20}$$

By observing the fact that $\mathcal{I}_s^{\mathbf{x}} = \text{supp}(\mathbf{x}^*)$ for all $\mathbf{x} \in \mathcal{B} \cap C_s$, let $\mathcal{S}_{\mathcal{L}}$ be a subspace defined by $\mathcal{I}_s^{\mathbf{x}}$. As f is strictly convex on any $\mathcal{S}_{\mathcal{J}}$, $f(\mathbf{x}) \geq f(\mathbf{x}^*)$ for all $\mathbf{x} \in \mathcal{B} \cap C_s$ such that $\mathbf{x}^*$ is a unique local minimizer of Problem (2) on $\mathcal{B} \cap C_s$. Further, in light of 20 we obtain that for any given $\mathbf{x} \in \mathcal{B} \cap C_s$,

$H_s(\mathbf{x}-\gamma\nabla f(\mathbf{x}))$ has a unique solution $\mathbf{y}$ with $\text{supp}(\mathbf{y})=\mathcal{L}$, and $\mathbf{y}_{\mathcal{L}}=\mathbf{x}_{\mathcal{L}}-\gamma_{\mathcal{L}}f(\mathbf{x})$. Hence, $\mathbf{y}\in\mathcal{S}_{\mathcal{L}}$. We claim that

$$\|\mathbf{y}-\mathbf{x}^*\|_2\leq\|\mathbf{x}-\mathbf{x}^*\|_2.$$

Towards this end, we see via $\mathbf{y}\in\mathcal{S}_{\mathcal{L}}$ and the previous argument that $f(\mathbf{y})\geq f(\mathbf{x}^*)$. Furthermore, since $\mathbf{y}_{\mathcal{L}}=\mathbf{x}_{\mathcal{L}}-\gamma_{\mathcal{L}}f(\mathbf{x})$, we deduce from Corollary 2 and $\mathcal{I}_s^{\mathbf{x}}=\text{supp}(\mathbf{x}^*)=\mathcal{L}$ that

$$\frac{\gamma}{2}\|\nabla_{\mathcal{L}}f(\mathbf{x})\|_2^2\leq f(\mathbf{x})-f(\mathbf{y}).$$

Due to the (strict) convexity of f on $\mathcal{S}_{\mathcal{L}}$ and $\text{supp}(\mathbf{x})=\text{supp}(\mathbf{x}^*)=\mathcal{S}_{\mathcal{L}}$ and one has the following:

$$f(\mathbf{x}^*)\geq f(\mathbf{x})+\langle\nabla f(\mathbf{x}),\mathbf{x}^*-\mathbf{x}\rangle=f(\mathbf{x})+\langle\nabla_{\mathcal{L}}f(\mathbf{x}),\mathbf{x}^*_{\mathcal{L}}-\mathbf{x}_{\mathcal{L}}\rangle$$

which yields $f(\mathbf{x})-f(\mathbf{x}^*)\leq\langle\nabla_{\mathcal{L}}f(\mathbf{x}),\mathbf{x}_{\mathcal{L}}-\mathbf{x}^*_{\mathcal{L}}\rangle$. We further have

$$0\leq f(\mathbf{y})-f(\mathbf{x}^*)\leq f(\mathbf{x})-f(\mathbf{x}^*)-\frac{\gamma}{2}\|\nabla_{\mathcal{L}}f(\mathbf{x})\|_2^2\leq\langle\nabla_{\mathcal{L}}f(\mathbf{x}),\mathbf{x}_{\mathcal{L}}-\mathbf{x}^*_{\mathcal{L}}\rangle-\frac{\gamma}{2}\|\nabla_{\mathcal{L}}f(\mathbf{x})\|_2^2.$$

Using these results, we obtain

$$\begin{aligned}0\leq\langle\nabla_{\mathcal{L}}f(\mathbf{x}),\mathbf{x}_{\mathcal{L}}-\mathbf{x}^*_{\mathcal{L}}\rangle-\frac{\gamma}{2}\|\nabla_{\mathcal{L}}f(\mathbf{x})\|_2^2&=\frac{1}{2\gamma}\Big(\|\mathbf{x}_{\mathcal{L}}-\mathbf{x}^*_{\mathcal{L}}\|_2^2-\|\mathbf{x}_{\mathcal{L}}-\nabla_{\mathcal{L}}f(\mathbf{x})-\mathbf{x}^*_{\mathcal{L}}\|_2^2\Big)\\&=\frac{1}{2\gamma}\Big(\|\mathbf{x}_{\mathcal{L}}-\mathbf{x}^*_{\mathcal{L}}\|_2^2-\|\mathbf{y}_{\mathcal{L}}-\mathbf{x}^*_{\mathcal{L}}\|_2^2\Big)\end{aligned}$$

This shows that $\|\mathbf{y}-\mathbf{x}^*\|_2\leq\|\mathbf{x}-\mathbf{x}^*\|_2$ thus the claim holds.

In view of the above claim, we deduce via induction that for any $\mathbf{x}^0\in\mathcal{B}\cap C_s$, $\mathbf{x}^k\in\mathcal{B}\cap C_s$ for all $k\in\mathbb{N}$. Hence, the IHT sequence $(\mathbf{x}^k)$ is contained in $\mathcal{B}$ and thus is bounded such that it has an accumulation point. As shown in Theorem 4, all accumulation points $\tilde{\mathbf{x}}$ of an IHT sequence are *HT-stable* and satisfy $\nabla_{\text{supp}(\tilde{\mathbf{x}})}f(\tilde{\mathbf{x}})=0$ so does any accumulation point $\hat{\mathbf{x}}$ of $(\mathbf{x}^k)$. Since $(\mathbf{x}^k)$ is contained in $\mathcal{B}\cap C_s$, $\mathcal{I}_s^{\mathbf{x}^k}=\text{supp}(\mathbf{x}^*)$ for all $k\geq 0$ and $\hat{\mathbf{x}}$ satisfies $\text{supp}(\hat{\mathbf{x}})=\mathcal{S}_{\mathcal{L}}$ and $\nabla_{\mathcal{L}}f(\hat{\mathbf{x}})=0$. Thus, $\hat{\mathbf{x}}_{\mathcal{L}}=\mathbf{x}^*_{\mathcal{L}}$; otherwise, due to strict convexity of f, one can write $f(\hat{\mathbf{x}})>f(\mathbf{x}^*)+\langle\nabla_{\mathcal{L}}f(\mathbf{x}^*),\hat{\mathbf{x}}_{\mathcal{L}}-\mathbf{x}^*_{\mathcal{L}}\rangle=f(\mathbf{x}^*)$ and $f(\mathbf{x}^*)>f(\hat{\mathbf{x}})+\langle\nabla_{\mathcal{L}}f(\hat{\mathbf{x}}),\mathbf{x}^*_{\mathcal{L}}-\hat{\mathbf{x}}_{\mathcal{L}}\rangle=f(\hat{\mathbf{x}})$ which are impossible. This shows that any IHT sequence $(\mathbf{x}^k)$ in $\mathcal{B}$ has exactly one accumulation point given by $\mathbf{x}^*$ and thus converges to $\mathbf{x}^*$.

To show Q-linearly convergence suppose that f is strongly convex on $\mathcal{S}_{\mathcal{J}}$ for all index subsets $\mathcal{J}$ with $|\mathcal{J}|$. Then there exists a positive constant $m_{\mathcal{J}}$ with $0<m_{\mathcal{J}}\leq L_s$ such that

$$f(\mathbf{y})\geq f(\mathbf{x})+\langle\nabla_{\mathcal{J}}f(\mathbf{x}),\mathbf{y}_{\mathcal{J}}-\mathbf{x}_{\mathcal{J}}\rangle+\frac{m_{\mathcal{J}}}{2}\|\mathbf{y}-\mathbf{x}\|_2^2,\quad\forall\mathbf{y},\mathbf{x}\in\mathcal{S}_{\mathcal{J}}$$

The prior argument shows that $\nabla_{\mathcal{J}}f(\mathbf{x}^*)=0$. Since f is strongly convex on $\mathcal{S}_{\mathcal{J}}$, we have $f(\mathbf{y})\geq f(\mathbf{x})+\frac{m_{\mathcal{J}}}{2}\|\mathbf{y}-\mathbf{x}\|_2^2$ for all $\mathbf{y}\in\mathcal{S}_{\mathcal{J}}$. It follows from the similar argument as before that for every $\mathbf{x}\in\mathcal{B}\cap C_s$, $H_s(\mathbf{x}-\gamma\nabla f(\mathbf{x}))$ has a unique solution $\mathbf{x}$ with $\text{supp}(\mathbf{y})=\mathcal{L}$ and $\mathbf{y}_{\mathcal{L}}=\mathbf{x}_{\mathcal{L}}-\gamma_{\mathcal{L}}f(\mathbf{x})$. Hence, $\mathbf{y}\in\mathcal{S}_{\mathcal{J}}$. As a result, we obtain, in view of $0<\frac{1}{\gamma}\leq L_s$,

$$\begin{aligned}\frac{m_{\mathcal{J}}}{2}\|\mathbf{y}-\mathbf{x}^*\|_2^2\leq f(\mathbf{y})-f(\mathbf{x}^*)&=\frac{1}{2\gamma}\Big(\|\mathbf{x}_{\mathcal{L}}-\mathbf{x}^*_{\mathcal{L}}\|_2^2-\|\mathbf{x}_{\mathcal{L}}-\nabla_{\mathcal{L}}f(\mathbf{x})-\mathbf{x}^*_{\mathcal{L}}\|_2^2\Big)\\&=\frac{L_s}{2}\Big(\|\mathbf{x}_{\mathcal{L}}-\mathbf{x}^*_{\mathcal{L}}\|_2^2-\|\mathbf{y}_{\mathcal{L}}-\mathbf{x}^*_{\mathcal{L}}\|_2^2\Big)\\&=\frac{L_s}{2}\Big(\|\mathbf{x}-\mathbf{x}^*\|_2^2-\|\mathbf{y}-\mathbf{x}^*\|_2^2\Big).\end{aligned}$$

This give rise to

$$\|\mathbf{y}-\mathbf{x}^*\|_2^2\leq\frac{L_s}{L_s+m_{\mathcal{J}}}\|\mathbf{x}-\mathbf{x}^*\|_2^2$$

Obviously, this implies that any IHT sequence $(\mathbf{x}^k)$ in $\mathcal{B}$ Q-linearly converges to $\mathbf{x}^*$. □