A. Technical Lemmas

The following lemma is a characterization of the co-coercivity of the objective function \( F(x) \). A similar result was obtained in Nguyen et al. (2014, Corollary 8) but we present a refined analysis which is essential for our purpose.

**Lemma 9.** For a given support set \( \Omega \), assume that the continuous function \( F(x) \) is \( M_{|\Omega|} \)-RSS and is \( m_K \)-RSC for some sparsity level \( K \). Then, for all vectors \( w \) and \( w' \) with \( |\text{supp} (w - w') \cup \Omega| \leq K \), we have

\[
\|\nabla_{\Omega} F(w') - \nabla_{\Omega} F(w)\|^2 \leq 2M_{|\Omega|} (F(w') - F(w) - \langle \nabla F(w), w' - w \rangle).
\]

**Proof.** We define an auxiliary function

\[
G(x) \overset{\text{def}}{=} F(x) - \langle \nabla F(w), x \rangle.
\]

For all vectors \( x \) and \( y \), we have

\[
\|\nabla G(x) - \nabla G(y)\| = \|\nabla F(x) - \nabla F(y)\| \leq M_{|\text{supp}(x - y)|} \|x - y\|,
\]

which is equivalent to

\[
G(x) - G(y) - \langle \nabla G(y), x - y \rangle \leq \frac{M_r}{2} \|x - y\|^2,
\]

where \( r = |\text{supp}(x - y)| \). On the other hand, due to the RSC property of \( F(x) \), we obtain

\[
G(x) - G(w) = F(x) - F(w) - \langle \nabla F(w), x - w \rangle \geq \frac{m_{|\text{supp}(x - w)|}}{2} \|x - w\|^2 \geq 0,
\]

provided that \( |\text{supp}(x - w)| \leq K \). For the given support set \( \Omega \), we pick \( x = w' - \frac{1}{M_{|\Omega|}} \nabla_{\Omega} G(w') \). Clearly, for such a choice of \( x \), we have \( \text{supp}(x - w) = \text{supp}(w - w') \cup \Omega \). Hence, by assuming that \( |\text{supp}(w - w') \cup \Omega| \) is not larger than \( K \), we get

\[
G(w) \leq G \left( w' - \frac{1}{M_{|\Omega|}} \nabla_{\Omega} G(w') \right) \overset{(\ref{eq:lemma9})}{\leq} G(w') + \left\langle \nabla G(w'), -\frac{1}{M_{|\Omega|}} \nabla_{\Omega} G(w') \right\rangle + \frac{1}{2M_{|\Omega|}} \|\nabla_{\Omega} G(w')\|^2
\]

\[
= G(w') - \frac{1}{2M_{|\Omega|}} \|\nabla_{\Omega} G(w')\|^2.
\]

Now expanding \( \nabla_{\Omega} G(w') \) and rearranging the terms give the desired result.

**Lemma 10** (Lemma 1 in Wang et al. (2016)). Let \( u \) and \( z \) be two distinct vectors and let \( W = \text{supp} (u) \cap \text{supp} (z) \). Also, let \( U \) be the support set of the top \( r \) (in magnitude) elements in \( u \). Then, the following holds for all \( r \geq 1 \):

\[
\langle u, z \rangle \leq \sqrt{\frac{|W|}{r}} \|u_U\| \cdot \|z_W\|.
\]

**Lemma 11.** Suppose that \( F(x) \) is \( m_K \)-restricted strongly convex and \( M_K \)-restricted smooth for some sparsity level \( K > 0 \). Then for all \( \eta > 0 \), all vectors \( x, x' \in \mathbb{R}^d \) and for any Hessian matrix \( H \) of \( F(x) \), we have

\[
|\langle x, (I - \eta H)x' \rangle| \leq \rho \|x\| \cdot \|x'\|, \quad \text{if } |\text{supp}(x) \cup \text{supp}(x')| \leq K,
\]

and

\[
\|(I - \eta H)x_S\| \leq \rho \|x\|, \quad \text{if } |S \cup \text{supp}(x)| \leq K,
\]

where

\[
\rho = \max \{ |\eta m_K - 1|, |\eta M_K - 1| \}.
\]
Proof. Since $H$ is a Hessian matrix, we always have a decomposition $H = A^T A$ for some matrix $A$. Denote $T = \text{supp}(x) \cup \text{supp}(y)$. By simple algebra, we have

$$|\langle x, (I - \eta H)x' \rangle| = |\langle x, x' \rangle - \eta \langle Ax, Ax' \rangle|$$

$$\leq |\langle x, x' \rangle - \eta \langle A_T x, A_T x' \rangle|$$

$$= |\langle x, (I - \eta A_T^T A_T)x' \rangle|$$

$$\leq \|I - \eta A_T^T A_T\| \cdot \|x'\|$$

$$\leq \max \{|\eta m_K - 1|, |\eta m_K - 1|\} \cdot \|x\| \cdot \|x'\|.$$ 

Here, $\zeta_1$ follows from the fact that $\text{supp}(x) \cup \text{supp}(y) = T$ and $\zeta_2$ holds because the RSC and RSS properties imply that the singular values of any Hessian matrix restricted on an $K$-sparse support set are lower and upper bounded by $m_K$ and $M_K$, respectively. For some index set $S$ subject to $|S \cup \text{supp}(x)| \leq K$, let $x' = (I - \eta H)x_S$. We immediately obtain

$$\|x'\|^2 = \langle x', (I - \eta H)x \rangle \leq \rho \|x'\| \cdot \|x\|,$$

indicating

$$\|((I - \eta H)x)_S\| \leq \rho \|x\|. \quad \square$$

Lemma 12. Suppose that $F(x)$ is $m_K$-restricted strongly convex and $M_K$-restricted smooth for some sparsity level $K > 0$. For all $\eta > 0$, all vectors $x, x' \in \mathbb{R}^d$ and support set $T$ such that $|\text{supp}(x - x') \cup T| \leq K$, the following holds:

$$\|(x - x' - \eta \nabla F(x) + \eta \nabla F(x'))_T\| \leq \rho \|x - x'\|$$

where $\rho$ is given in Lemma 11.

Proof. In fact, for any two vectors $x$ and $x'$, there always exists a quantity $\theta \in [0, 1]$, such that

$$\nabla F(x) - \nabla F(x') = \nabla^2 F(\theta x + (1 - \theta)x') (x - x').$$

Let $H = \nabla^2 F(\theta x + (1 - \theta)x')$. We write

$$\|(x - x' - \eta \nabla F(x) + \eta \nabla F(x'))_T\| = \|(x - x' - \eta H(x - x'))_T\|$$

$$= \|((I - \eta H)(x - x'))_T\|$$

$$\leq \rho \|x - x'\|,$$

where the last inequality applies Lemma 11. \quad \square

Lemma 13. Suppose that $x$ is a $k$-sparse vector and let $b = x - \eta \nabla F(x)$. Let $T$ be the support set that contains the $k$ largest absolute values of $b$. Assume that the function $F(x)$ is $M_{2k}$-restricted smooth, then we have the following:

$$F(b_T) \leq F(x) - \frac{1 - \eta M_{2k}}{2\eta} \|b_T - x\|^2.$$

Proof. The RSS condition implies that

$$F(b_T) - F(x) \leq \langle \nabla F(x), b_T - x \rangle + \frac{M_{2k}}{2} \|b_T - x\|^2$$

$$\leq - \frac{1}{2\eta} \|b_T - x\|^2 + \frac{M_{2k}}{2} \|b_T - x\|^2,$$
Support Recovery of Hard Thresholding Pursuit

where the second inequality is due to the fact that
\[
\|b_T - b\|^2 = \|b_T - x + \eta \nabla F(x)\|^2 \\
\leq \|x - x + \eta \nabla F(x)\|^2 \\
= \|\eta \nabla F(x)\|^2,
\]

implying
\[
2\eta \langle \nabla F(x), b_T - x \rangle \leq -\|b_T - x\|^2.
\]

Lemma 14. Suppose that \( F(x) \) is \( m_K \)-RSC. Then for any vectors \( x \) and \( x' \) with \( \|x - x'\|_0 \leq K \), the following holds:
\[
\|x - x'\| \leq \sqrt{\frac{2 \max\{F(x) - F(x'), 0\}}{m_K}} + \frac{2 \|\nabla_T F(x')\|}{m_K}
\]
where \( T = \text{supp} (x - x') \).

Proof. The RSC property immediately implies
\[
F(x) - F(x') \geq \langle \nabla F(x'), x - x' \rangle + \frac{mK}{2} \|x - x'\|^2 \\
\geq -\|\nabla_T F(x')\| \cdot \|x - x'\| + \frac{mK}{2} \|x - x'\|^2.
\]
Discussing the sign of \( F(x) - F(x') \) and solving the above quadratic inequality completes the proof.

Lemma 15. Assume that \( F(x) \) is \( m_{k+s} \)-RSC and \( M_{2k} \)-RSS. Suppose that for all \( t \geq 0 \), \( x^t \) is \( k \)-sparse and the following holds:
\[
F(x^{t+1}) - F(\bar{x}) \leq \mu (F(x^t) - F(\bar{x})) + \tau,
\]
where \( 0 < \mu < 1 \), \( \tau \geq 0 \) and \( \bar{x} \) is an arbitrary \( s \)-sparse signal. Then,
\[
\|x^t - \bar{x}\| \leq \sqrt{\frac{2M}{m} (\sqrt{\mu})^t \|x^0 - \bar{x}\|} + \frac{3}{m} \|\nabla_{k+s} F(\bar{x})\| + \sqrt{\frac{2\tau}{m(1-\mu)}}.
\]

Proof. The RSS property implies that
\[
F(x^0) - F(\bar{x}) \leq \langle \nabla F(\bar{x}), x^0 - \bar{x} \rangle + \frac{M}{2} \|x^0 - \bar{x}\|^2 \\
\leq \frac{M}{2} \|x^0 - \bar{x}\|^2 + \frac{1}{2M} \|\nabla_{k+s} F(\bar{x})\|^2 + \frac{M}{2} \|x^0 - \bar{x}\|^2 \\
\leq M \|x^0 - \bar{x}\|^2 + \frac{1}{2M} \|\nabla_{k+s} F(\bar{x})\|^2.
\]
Hence,
\[
F(x^t) - F(\bar{x}) \leq \mu^t M \|x^0 - \bar{x}\|^2 + \frac{1}{2M} \|\nabla_{k+s} F(\bar{x})\|^2 + \frac{\tau}{1-\mu}.
\]
By Lemma 14, we have
\[
\|x' - \tilde{x}\| \leq \sqrt{\frac{2M}{m}} (\mu M \|x^0 - \tilde{x}\|^2 + \frac{\|\nabla k_+ s F(\tilde{x})\|^2}{2M} + \frac{\tau}{1-\mu} + \frac{2}{m} \|\nabla k_+ s F(\tilde{x})\|)
\]
\[
\leq \sqrt{\frac{2M}{m}} (\mu M \|x^0 - \tilde{x}\|^2 + \frac{\|\nabla k_+ s F(\tilde{x})\|^2}{2M} + \frac{2\tau}{m(1-\mu)})
\]
\[
\leq \sqrt{\frac{2M}{m}} (\mu M \|x^0 - \tilde{x}\|^2 + \frac{3}{m} \|\nabla k_+ s F(\tilde{x})\| + \frac{2\tau}{m(1-\mu)}).
\]

Lemma 16. Let \(\tilde{x} \in \mathbb{R}^d\) be an \(s\)-sparse vector supported on \(S\). For a \(k\)-sparse vector \(x\) supported on \(Q\) with \(k \geq s\), let
\(b = x - \eta \nabla F(x)\) and let \(T = \text{supp}(b, k)\). Suppose that the function \(F(x)\) is \(m_{2k+s}\)-RSC and \(M_{2k+s}\)-RSS. Then we have
\[
\|x_{S \setminus T}\| \leq \nu \rho \|x - \tilde{x}\| + \nu \eta \|\nabla T_\Delta S F(\tilde{x})\|
\]
where \(\nu = \sqrt{1 + s/k}\) and \(\rho\) is given by Lemma 11.

Proof. We note the fact that the support sets \(T \setminus S\) and \(S \setminus T\) are disjoint. Moreover, the set \(T \setminus S\) contains \(|T\setminus S|\) number of top \(|T|\) elements of \(b\). Hence, we have
\[
\frac{1}{|T \setminus S|} \|b_{T \setminus S}\|^2 \geq \frac{1}{|S \setminus T|} \|b_{S \setminus T}\|^2.
\]
That is,
\[
\|b_{T \setminus S}\| \geq \sqrt{\frac{|T \setminus S|}{|S \setminus T|}} \|b_{S \setminus T}\| = \sqrt{\frac{k - |T \cap S|}{s - |T \cap S|}} \|b_{S \setminus T}\| \geq \sqrt{\frac{k}{s}} \|b_{S \setminus T}\|.
\]
Note that the above holds also for \(T = S\). Since \(\tilde{x}\) is supported on \(S\), the left hand side reads as
\[
\|b_{T \setminus S}\| = \left\| (x - \tilde{x} - \eta \nabla F(x))_{T \setminus S} \right\|
\]
while the right hand side reads as
\[
\|b_{S \setminus T}\| = \left\| (x - \tilde{x} - \eta \nabla F(x))_{S \setminus T} + \tilde{x}_{S \setminus T} \right\|
\]
\[
\geq \left\| \tilde{x}_{S \setminus T} \right\| - \left\| (x - \tilde{x} - \eta \nabla F(x))_{S \setminus T} \right\|
\]
Denote \(\nu = \sqrt{1 + s/k}\). In this way, we arrive at
\[
\|\tilde{x}_{S \setminus T}\| \leq \sqrt{s} \left\| (x - \tilde{x} - \eta \nabla F(x))_{T \setminus S} \right\| + \left\| (x - \tilde{x} - \eta \nabla F(x))_{S \setminus T} \right\|
\]
\[
\leq \nu \left\| (x - \tilde{x} - \eta \nabla F(x))_{T \Delta S} \right\|
\]
\[
\leq \nu \left\| (x - \tilde{x} - \eta \nabla F(x) + \eta \nabla F(\tilde{x}))_{T \Delta S} \right\| + \nu \eta \|\nabla T_\Delta S F(\tilde{x})\|
\]
\[
\leq \nu \left\| (x - \tilde{x} - \eta \nabla F(x) + \eta \nabla F(\tilde{x}))_{T \Delta Q_{\Delta S}} \right\| + \nu \eta \|\nabla T_\Delta S F(\tilde{x})\|
\]
\[
\leq \nu \rho_{2k+s} \|x - \tilde{x}\| + \nu \eta \|\nabla T_\Delta S F(\tilde{x})\|
\]
where the second inequality follows from the fact that \(ax + by \leq \sqrt{a^2 + b^2} \sqrt{x^2 + y^2}\) and we applied Lemma 12 for the last inequality. \(\square\)
Lemma 17. Consider the HTP algorithm with exact solutions. Assume (A1). Then

\[ \| \nabla_{S^{t+1} \setminus S^t} F(x^t) \|^2 \geq 2m \zeta \left( F(x^t) - F(\bar{x}) \right), \]

where

\[ \zeta = \frac{|S^{t+1} \setminus S^t|}{|S^{t+1} \setminus S^t| + |S \setminus S^t|}. \]

Proof. The lemma holds clearly for either \( S^{t+1} = S^t \) or \( F(x^t) \leq F(\bar{x}) \). Hence, in the following we only prove the result by assuming \( S^{t+1} \neq S^t \) and \( F(x^t) > F(\bar{x}) \). Due to the RSC property, we have

\[ F(\bar{x}) - F(x^t) - \langle \nabla F(x^t), \bar{x} - x^t \rangle \geq \frac{m_{k+s}}{2} \| \bar{x} - x^t \|^2, \]

which implies

\[ \langle \nabla F(x^t), -\bar{x} \rangle \geq \frac{m_{k+s}}{2} \| \bar{x} - x^t \|^2 + F(x^t) - F(\bar{x}) \]

\[ \geq \sqrt{2m_{k+s}} \| \bar{x} - x^t \| \sqrt{F(x^t) - F(\bar{x})}. \]

By invoking Lemma 10 with \( u = \nabla F(x^t) \) and \( z = -\bar{x} \) therein, we have

\[ \langle \nabla F(x^t), -\bar{x} \rangle \leq \sqrt{\frac{|S \setminus S^t|}{|S^{t+1} \setminus S^t|} + 1} \| \nabla_{S^{t+1} \setminus S^t} F(x^t) \| \cdot \| \bar{x} \|_{S^t} \]

\[ \leq \sqrt{\frac{|S \setminus S^t|}{|S^{t+1} \setminus S^t|} + 1} \| \nabla_{S^{t+1} \setminus S^t} F(x^t) \| \cdot \| \bar{x} - x^t \|. \]

It is worth mentioning that the first inequality above holds because \( \nabla F(x^t) \) is supported on \( S^t \) and \( S^{t+1} \setminus S^t \) contains the \( |S^{t+1} \setminus S^t| \) number of largest (in magnitude) elements of \( \nabla F(x^t) \). Therefore, we obtain the result. \( \square \)

B. Proofs for Section 2

B.1. Proof for Prop. 1

Proof. Due to the RSS property, we have

\[ F(b_{S^{t+1}}^{t+1}) - F(x^t) \leq \langle \nabla F(x^t), b_{S^{t+1}}^{t+1} - x^t \rangle + \frac{M}{2} \| b_{S^{t+1}}^{t+1} - x^t \|^2 \]

\[ \overset{\zeta_1}{=} \langle \nabla_{S^{t+1} \setminus S^t} F(x^t), b_{S^{t+1} \setminus S^t}^{t+1} \rangle + \frac{M}{2} \left( \| b_{S^{t+1} \setminus S^t}^{t+1} \| \right)^2 \]

\[ + \| b_{S^{t+1} \setminus S^t}^{t+1} - x_{S^{t+1} \setminus S^t}^t \|^2 + \| x_{S^t \setminus S^{t+1}} \|^2 \]

\[ \overset{\zeta_2}{=} \langle \nabla_{S^{t+1} \setminus S^t} F(x^t), b_{S^{t+1} \setminus S^t}^{t+1} \rangle + M \| b_{S^{t+1} \setminus S^t}^{t+1} \|^2 \]

\[ \overset{\zeta_3}{=} -\eta(1 - \eta M) \| \nabla_{S^{t+1} \setminus S^t} F(x^t) \|^2. \]

Above, we observe that \( \nabla F(x^t) \) is supported on \( S^t \) and we simply decompose the support set \( S^{t+1} \cup S^t \) into three mutually disjoint sets, and hence \( \zeta_1 \) holds. To see why \( \zeta_2 \) holds, we note that for any set \( \Omega \subset S^t, b_{\Omega}^{t+1} = x_{\Omega}^t \). Hence, \( b_{S^{t+1} \setminus S^t}^{t+1} = x_{S^{t+1} \setminus S^t}^t \). Moreover, since \( x_{S^t \setminus S^{t+1}}^t = b_{S^t \setminus S^{t+1}}^{t+1} \) and any element in \( b_{S^{t+1} \setminus S^t}^{t+1} \) is not larger than that in \( b_{S^{t+1} \setminus S^t}^{t+1} \) (recall that \( S^{t+1} \) is obtained by hard thresholding), we have \( \| x_{S^t \setminus S^{t+1}}^t \| \leq \| b_{S^{t+1} \setminus S^t}^{t+1} \| \) where we use the fact that \( |S^t \setminus S^{t+1}| = |S^{t+1} \setminus S^t| \). Therefore, \( \zeta_2 \) holds. Finally, we write \( b_{S^{t+1} \setminus S^t}^{t+1} = -\eta \nabla_{S^{t+1} \setminus S^t} F(x^t) \) and obtain \( \zeta_3 \).
Since $x^{t+1}$ is a minimizer of $F(x)$ over the support set $S^{t+1}$, it immediately follows that
\begin{equation}
F(x^{t+1}) - F(x^t) \leq F(b_{S^{t+1}}^t - F(x^t) \leq -\eta(1 - \eta M) \|\nabla_{S^{t+1}\setminus S^t} F(x^t)\|^2.
\end{equation}

Now we invoke Lemma 17 and pick $\eta \leq 1/M$,
\begin{equation}
F(x^{t+1}) - F(x^t) \leq \eta (\eta M - 1) \cdot \frac{2m}{1 + s} (F(x^t) - F(\bar{x})),
\end{equation}
which gives
\begin{equation}
F(x^{t+1}) - F(\bar{x}) \leq \beta (F(x^t) - F(\bar{x})),
\end{equation}
where $\beta = 1 - \frac{2m\eta(1 - \eta M)}{1 + s}$.

**B.2. Proof for Prop. 2**

*Proof.* This is a direct result by combining Prop. 1 and Lemma 15.

**B.3. Proof for Lemma 3**

*Proof.* Let $x^*_t = \arg\min_{\supp(x) \subset S^t} F(x)$. Since $x^t$ and $x^*_t$ are both supported on $S^t$, we apply Lemma 9 and obtain
\begin{equation}
\|\nabla_{S^t} F(x^t)\|^2 = \|\nabla_{S^t} F(x^t) - \nabla_{S^t} F(x^*_t)\|^2 \\
\leq 2M \left( F(x^t) - F(x^*_t) - \langle \nabla F(x^*_t), x^t - x^*_t \rangle \right) \\
\leq 2M \epsilon.
\end{equation}

Above, the second inequality uses the fact that $\nabla_{S^t} F(x^*_t) = 0$ and $F(x^t) \leq F(x^*_t) + \epsilon$.

**B.4. Proof for Prop. 4**

*Proof.* We have by Lemma 16 that
\begin{equation}
\|\bar{x}_{S^{t+1}}\| \leq \sqrt{2} \rho \|x^t - \bar{x}\| + \frac{2}{m} \|\nabla_{k+s} F(\bar{x})\|,
\end{equation}
where $\rho = 1 - \eta m$. On the other hand, Lemma 18 together with Lemma 3 shows that
\begin{equation}
\|x^{t+1} - \bar{x}\| \leq \kappa \|\bar{x}_{S^{t+1}}\| + \frac{1}{m} \|\nabla_{k} F(\bar{x})\| + \frac{1}{m} \sqrt{2M \epsilon}.
\end{equation}

Therefore,
\begin{equation}
\|x^{t+1} - \bar{x}\| \leq \sqrt{2} \kappa \rho \|x^t - \bar{x}\| + \frac{3\kappa}{m} \|\nabla_{k+s} F(\bar{x})\| + \frac{\sqrt{2M \epsilon}}{m}.
\end{equation}

We need to ensure
\begin{equation}
\sqrt{2} \kappa (1 - \eta m) < 1.
\end{equation}

Let $\eta = \eta'/M$ with $\eta' < 1$. Then, the above holds provided that
\begin{equation}
\kappa < 1 + \frac{1}{\sqrt{2}} \text{ and } \eta' > \kappa - \frac{1}{\sqrt{2}}.
\end{equation}

By induction and picking proper $\eta'$ to make $\sqrt{2\kappa(1 - \eta m)} < \sqrt{2}/4$, we have
\begin{equation}
\|x^t - \bar{x}\| \leq (\sqrt{2}(\kappa - \eta'))^t \|x^0 - \bar{x}\| + \frac{6\kappa}{m} \|\nabla_{k+s} F(\bar{x})\| + \frac{4\sqrt{M \epsilon}}{m}.
\end{equation}
B.5. Proof for Prop. 5

Proof. Our proof in this part is inspired by Yuan et al. (2016). Let \( \bar{x}_t = \arg\min_{\supp(x) \subseteq S_t} F(x) \). Then

\[
F(x^t) - F(x^{t-1}) \leq F(x^t) - F(x^{t-1}) + \epsilon \\
\leq F(b^t_{S_t}) - F(x^{t-1}) + \epsilon \\
\leq -\frac{1 - \eta M}{2\eta} \|b^t_{S_t} - x^{t-1}\|^2 + \epsilon,
\]

where the last inequality follows from Lemma 13. Now we bound the term \( \|b^t_{S_t} - x^{t-1}\|^2 \). Note that \( x^{t-1} \) is supported on \( S^{t-1} \). Hence,

\[
\|b^t_{S_t} - x^{t-1}\|^2 = \|x^{t-1}_{S_t \cap S^{t-1}} - \eta \nabla_{S_t} F(x^{t-1}) - x^{t-1}\|^2 \\
= \|x^{t-1}_{S_t \cap S^{t-1}} - \eta \nabla_{S_t} F(x^{t-1})\|^2 \\
= \|x^{t-1}_{S_t \cap S^{t-1}}\|^2 + \eta^2 \|\nabla_{S_t} F(x^{t-1})\|^2 \\
\geq \eta^2 \|\nabla_{S_t \setminus S^{t-1}} F(x^{t-1})\|^2.
\]

We thus have

\[
F(x^t) - F(x^{t-1}) \leq -\frac{(1 - \eta M)\eta}{2} \|\nabla_{S_t \setminus S^{t-1}} F(x^{t-1})\|^2 + \epsilon.
\]

Denote \( \xi = \|\nabla_{S^{t-1}} F(x^{t-1})\| \). We claim that

\[
\|\nabla_{S_t \setminus S^{t-1}} F(x^{t-1})\|^2 \geq m (F(x^{t-1}) - F(\bar{x})) - 2\xi^2,
\]

which, combined with Lemma 3, immediately shows

\[
F(x^t) - F(x^{t-1}) \leq -\frac{(1 - \eta M)\eta m}{2} (F(x^{t-1}) - F(\bar{x})) + 2\epsilon.
\]

Using Lemma 15 completes the proof.

To show (9), we consider two exhaustive cases: \(|S^t \setminus S^{t-1}| \geq s\) and \(|S^t \setminus S^{t-1}| < s\), and prove that (9) holds for both cases.

Case I. \(|S^t \setminus S^{t-1}| \geq s\). Due to the RSC property, we have

\[
\frac{m}{2} \|\bar{x} - x^{t-1}\|^2 \\
\leq F(\bar{x}) - F(x^{t-1}) - \langle \nabla F(x^{t-1}), \bar{x} - x^{t-1} \rangle \\
\leq F(\bar{x}) - F(x^{t-1}) + \frac{m}{2} \|\bar{x} - x^{t-1}\|^2 + \frac{1}{2m} \|\nabla_{S_t \setminus S^{t-1}} F(x^{t-1})\|^2 \\
= F(\bar{x}) - F(x^{t-1}) + \frac{m}{2} \|\bar{x} - x^{t-1}\|^2 + \frac{1}{2m} \|\nabla_{S_t \setminus S^{t-1}} F(x^{t-1})\|^2 + \frac{1}{2m} \|\nabla_{S_t \setminus S^{t-1}} F(x^{t-1})\|^2 \\
= F(\bar{x}) - F(x^{t-1}) + \frac{m}{2} \|\bar{x} - x^{t-1}\|^2 + \frac{1}{2m} \|\nabla_{S_t \setminus S^{t-1}} F(x^{t-1})\|^2 + \frac{1}{2m} \xi^2.
\]

Therefore, we get

\[
\|\nabla_{S_t \setminus S^{t-1}} F(x^{t-1})\|^2 \geq 2m (F(x^{t-1}) - F(\bar{x})) - \xi^2.
\]

Since \( S^t \) contains the \( k \) largest absolute values of \( b^t \), and \(|S^t \setminus S^{t-1}| \geq s \geq |S \setminus S^{t-1}|\), we have

\[
\|b^t_{S^t \setminus S^{t-1}}\|^2 \geq \|b^t_{S^t \setminus S^{t-1}}\|^2,
\]

Support Recovery of Hard Thresholding Pursuit
which immediately implies (9) by noting the fact that $b_{S^t \setminus S^{t-1}} = -\eta \nabla_{S^t \setminus S^{t-1}} F(x^{t-1})$ and $b_{S^t \setminus S^{t-1}} = -\eta \nabla_{S^t \setminus S^{t-1}} F(x^{t-1})$.

Case II. $|S^t \setminus S^{t-1}| < s$. Again, we use the RSC property to obtain

$$
\frac{m}{2} \| \bar{x} - x^{t-1} \|^2 
\leq F(\bar{x}) - F(x^{t-1}) - \langle \nabla F(x^{t-1}), \bar{x} - x^{t-1} \rangle 
\leq F(\bar{x}) - F(x^{t-1}) + \frac{m}{4} \| \bar{x} - x^{t-1} \|^2 + \frac{1}{m} \| \nabla_{S^t \setminus S^{t-1}} F(x^{t-1}) \|^2 
= F(\bar{x}) - F(x^{t-1}) + \frac{m}{4} \| \bar{x} - x^{t-1} \|^2 + \frac{1}{m} \| \nabla_{S^t \setminus S^{t-1}} F(x^{t-1}) \|^2 + \frac{1}{m} \xi^2 
= F(\bar{x}) - F(x^{t-1}) + \frac{m}{4} \| \bar{x} - x^{t-1} \|^2 + \frac{1}{m} \| \nabla_{(S^t \setminus S^{t-1}) \cap S} F(x^{t-1}) \|^2 
+ \frac{1}{m} \| \nabla_{(S^t \setminus S^{t-1}) \cap S} F(x^{t-1}) \|^2 + \frac{1}{m} \xi^2 
\leq F(\bar{x}) - F(x^{t-1}) + \frac{m}{4} \| \bar{x} - x^{t-1} \|^2 + \frac{1}{m} \| \nabla_{(S^t \setminus S^{t-1}) \cap S} F(x^{t-1}) \|^2 
+ \frac{1}{m} \| \nabla_{S^t \setminus S^{t-1}} F(x^{t-1}) \|^2 + \frac{1}{m} \xi^2. 
$$

We consider the term $\| \nabla_{S^t \setminus S^{t-1}} F(x^{t-1}) \|^2$ above. Actually, we have

$$
b_{S^t \setminus S^{t-1}} = -\eta \nabla_{S^t \setminus S^{t-1}} F(x^{t-1}).
$$

Since $S^t$ contains the $k$ largest absolute values of $b^t$, we know that any component in $b^t_{\Omega}$ is not larger than that in $b^t_{S^t}$ subject to $\Omega \cap S^t = \emptyset$. In particular,

$$
\left\| b_{S^t \setminus S^{t-1}} \right\|_{S^t \setminus S^{t-1}}^2 \leq \left\| b_{S^t \setminus S^{t-1}} \right\|_{S^t \setminus S^{t-1}}^2.
$$

Note that $|S^t \setminus S^{t-1}| < s$ implies $|(S^t \cap S^{t-1}) \setminus S| \geq k - 2s$. Therefore,

$$
\eta^2 \| \nabla_{S^t \setminus S^{t-1}} F(x^{t-1}) \|^2 
\leq \frac{s}{k - 2s} \left\| x^{t-1} \right\|_{(S^t \cap S^{t-1}) \setminus S} - \eta \| \nabla_{S^t \setminus S^{t-1}} F(x^{t-1}) \|^2 
\leq \frac{2s}{k - 2s} \left\| x^{t-1} \right\|_{(S^t \cap S^{t-1}) \setminus S}^2 + \frac{2s \eta^2}{k - 2s} \| x^{t-1} \|^2 
= \frac{2s}{k - 2s} \left\| x^{t-1} - \bar{x} \right\|_{(S^t \cap S^{t-1}) \setminus S}^2 + \frac{2s \eta^2}{k - 2s} \| x^{t-1} \|^2 
\leq \frac{2s}{k - 2s} \left\| x^{t-1} - \bar{x} \right\|^2 + \frac{2s \eta^2}{k - 2s} \xi^2.
$$

Plugging the above into (10), we obtain

$$
\frac{m}{2} \| \bar{x} - x^{t-1} \|^2 
\leq F(\bar{x}) - F(x^{t-1}) + \frac{m}{4} \| \bar{x} - x^{t-1} \|^2 + \frac{2s}{k - 2s} \| x^{t-1} \|^2 
+ \frac{1}{m} \| \nabla_{S^t \setminus S^{t-1}} F(x^{t-1}) \|^2 + \frac{1}{m} \left( \frac{2s}{k - 2s} + 1 \right) \xi^2.
$$

Picking $k \geq 2s + \frac{8s}{\eta^2 m}$ gives

$$
\frac{m}{2} \| \bar{x} - x^{t-1} \|^2 
\leq F(\bar{x}) - F(x^{t-1}) + \frac{m}{2} \| \bar{x} - x^{t-1} \|^2 
+ \frac{1}{m} \| \nabla_{S^t \setminus S^{t-1}} F(x^{t-1}) \|^2 + \left( \frac{\eta^2 m}{4} + \frac{1}{m} \right) \xi^2.
$$

Since $\eta < 1/M$, $\frac{\eta^2 m^2}{4} + 1 < 2$. Therefore, by re-arranging the above inequality, we prove the claim (9).
C. Proofs for Section 3

The following result holds for all $F(x)$.

**Lemma 18.** Assume (A1) and (A2). For any $k$-sparse vector $x$ and $s$-sparse vector $\bar{x}$, we have

$$\|x - \bar{x}\| \leq \kappa \|\bar{x}\| + \frac{1}{m} \|\nabla_T F(x) - \nabla_T F(\bar{x})\|,$$

where $T$ is the support set of $x$.

**Proof.**

$$\|(x - \bar{x})_T\|^2 = \|x - \bar{x} - \tau \nabla F(x) + \tau \nabla F(\bar{x}), (x - \bar{x})_T\| + \tau \|\nabla F(x) - \nabla F(\bar{x}), (x - \bar{x})_T\|$$

$$\leq \|\|x - \bar{x} - \tau \nabla F(x) + \tau \nabla F(\bar{x})\| \cdot \|(x - \bar{x})_T\| + \|\|\nabla F(x) - \nabla F(\bar{x})\| \cdot \|(x - \bar{x})_T\|$$

$$\leq \|x - \bar{x} - \tau \nabla_{TS} F(x) + \tau \nabla_{TS} F(\bar{x})\| \cdot \|(x - \bar{x})_T\| + \|\nabla F(x) - \nabla F(\bar{x})\| \cdot \|(x - \bar{x})_T\|$$

Dividing both sides by $\|(x - \bar{x})_T\|$ gives

$$\|(x - \bar{x})_T\| \leq \rho \|x - \bar{x}\| + \tau \|\nabla_T F(x) - \nabla_T F(\bar{x})\|.$$

On the other hand,

$$\|x - \bar{x}\| \leq \|(x - \bar{x})_T\| + \|(x - \bar{x})_{\bar{T}}\|$$

$$\leq \rho \|x - \bar{x}\| + \tau \|\nabla_T F(x) - \nabla_T F(\bar{x})\| + \|\bar{x}_{\bar{T}}\|.$$

Hence, we have

$$\|x - \bar{x}\| \leq \frac{1}{1 - \rho} \|\bar{x}\| + \frac{\tau}{1 - \rho} \|\nabla_T F(x) - \nabla_T F(\bar{x})\|.$$

Picking $\tau = 1/M$ completes the proof.

In view of the exact (HTP3), we have

$$\|x^t - \bar{x}\| \leq \kappa \|\bar{x}\| + \frac{1}{m} \|\nabla_k F(\bar{x})\|.$$  \hspace{1cm} (11)

Now we present the crucial lemma. It is inspired by Bouchot et al. (2016) but we show a more general result.

**Lemma 19.** Consider the HTP algorithm. Assume (A1) and (A2). Further assume that the sequence of $\{x^t\}_{t \geq 0}$ satisfies

$$\|x^t - \bar{x}\| \leq \alpha \cdot \beta^t \|x^0 - \bar{x}\| + \phi,$$

$$\|x^t - \bar{x}\| \leq \gamma \|\bar{x}_{\bar{T}}\| + \psi,$$

for positive $\alpha, \phi, \gamma, \psi$ and $0 < \beta < 1$. Suppose that at the $n$-th iteration ($n \geq 0$), $S^n$ contains the indices of top $p$ (in magnitude) elements of $\bar{x}$. Then, for any integer $1 \leq q \leq s - p$, there exists an integer $r \geq 1$ determined by

$$\sqrt{2} |\bar{x}_{p+q}| > \alpha \gamma \cdot \beta^{r-1} \|\bar{x}_{\{p+1, \ldots, s\}}\| + \theta$$

where

$$\theta = \alpha \psi + \phi + \frac{1}{m} \|\nabla_2 F(\bar{x})\|,$$

such that $S^{n+r}$ contains the indices of top $p + q$ elements of $\bar{x}$ provided that $\theta \leq \sqrt{2} \lambda \bar{x}_{\text{min}}$ for some $\lambda \in (0, 1)$. 


Moreover, on one hand, for any $j \in [p + q]$,
\[
|b_{j}^{n+r}| = \left| \left( x_{n+r-1} - \eta \nabla F(x_{n+r-1}) \right)_{j} \right| \\
\geq |\bar{x}_{j}| - \left| \left( x_{n+r-1} - \bar{x} - \eta \nabla F(x_{n+r-1}) \right)_{j} \right| \\
\geq |\bar{x}_{p+q}| - \left( x_{n+r-1} - \bar{x} - \eta \nabla F(x_{n+r-1}) \right)_{j} .
\]

On the other hand, for all $i \in \mathcal{S}$,
\[
|b_{i}^{n+r}| = \left| \left( x_{n+r-1} - \bar{x} - \eta \nabla F(x_{n+r-1}) \right)_{i} \right| .
\]

Hence, we know that to guarantee (12), it suffices to ensure for all $j \in [p + q]$ and $i \in \mathcal{S}$ that
\[
|\bar{x}_{p+q}| > \left( x_{n+r-1} - \bar{x} - \eta \nabla F(x_{n+r-1}) \right)_{j} + \left| \left( x_{n+r-1} - \bar{x} - \eta \nabla F(x_{n+r-1}) \right)_{i} \right| .
\]

Note that the right-hand side is upper bounded as follows:
\[
\frac{1}{\sqrt{2}} \left| (x_{n+r-1} - \bar{x} - \eta \nabla F(x_{n+r-1}))_{j} \right| + \frac{1}{\sqrt{2}} \left| (x_{n+r-1} - \bar{x} - \eta \nabla F(x_{n+r-1}))_{i} \right| \\
\leq \left\| (x_{n+r-1} - \bar{x} - \eta \nabla F(x_{n+r-1}))_{\{j, i\}} \right\| \\
\leq \left\| (x_{n+r-1} - \bar{x} - \eta \nabla F(x_{n+r-1}) + \eta \nabla F(\bar{x}))_{\{j, i\}} \right\| + \eta \left\| \nabla (j, i) F(\bar{x}) \right\| \\
\leq \rho \left\| x_{n+r-1} - \bar{x} \right\| + \eta \left\| \nabla_{2} F(\bar{x}) \right\| \\
\leq \rho \alpha \cdot \beta^{r-1} \left\| x_{n} - \bar{x} \right\| + \rho \phi + \eta \left\| \nabla_{2} F(\bar{x}) \right\| .
\]

Moreover,
\[
\left\| x_{n} - \bar{x} \right\| \leq \gamma \left\| x_{\mathcal{S}} \right\| + \psi \leq \gamma \left\| \bar{x}_{\mathcal{S}} \right\| + \psi = \gamma \left\| \bar{x}_{\{p+1, \ldots, s\}} \right\| + \psi .
\]

Put all together, we have
\[
\frac{1}{\sqrt{2}} \left| (x_{n+r-1} - \bar{x} - \eta \nabla F(x_{n+r-1}))_{j} \right| + \frac{1}{\sqrt{2}} \left| (x_{n+r-1} - \bar{x} - \eta \nabla F(x_{n+r-1}))_{i} \right| \\
\leq \rho \alpha \gamma \cdot \beta^{r-1} \left\| \bar{x}_{\{p+1, \ldots, s\}} \right\| + \rho \alpha \psi + \rho \phi + \eta \left\| \nabla_{2} F(\bar{x}) \right\| \\
\leq \alpha \gamma \cdot \beta^{r-1} \left\| \bar{x}_{\{p+1, \ldots, s\}} \right\| + \alpha \psi + \phi + \frac{1}{m} \left\| \nabla_{2} F(\bar{x}) \right\| .
\]

Therefore, when
\[
\sqrt{2} |\bar{x}_{p+q}| > \alpha \gamma \cdot \beta^{r-1} \left\| \bar{x}_{\{p+1, \ldots, s\}} \right\| + \alpha \psi + \phi + \frac{1}{m} \left\| \nabla_{2} F(\bar{x}) \right\| ,
\]
we always have (12). Note that the above holds as far as $\alpha \psi + \phi + \frac{1}{m} \left\| \nabla_{2} F(\bar{x}) \right\|$ is strictly smaller than $\sqrt{2} |\bar{x}_{s}|$. \[\Box\]

With Lemma 19, we show the following general theorem.

**Theorem 20.** Assume same conditions as in Lemma 19. Then HTP successfully identifies the support of $\bar{x}$ using $\left( \frac{\log 2}{2 \log(1/\rho)} + \frac{\log(\alpha \gamma (1 - \lambda))}{\log(1/\rho)} + 2 \right)$ s number of iterations.
Proof. Without loss of generality, we presume that the elements in \( \bar{x} \) are in descending order by their magnitude, i.e., \( |\bar{x}_1| \geq |\bar{x}_2| \geq \cdots \geq |\bar{x}_s| \). We partition the support set \( [s] \) into \( K \) folds \( S_1, S_2, \ldots, S_K \), where each \( S_i \) is defined as follows:

\[
S_i = \{s_{i-1} + 1, \ldots, s_i\}, \; \forall 1 \leq i \leq K.
\]

Here, \( s_0 = 0 \) and for all \( 1 \leq i \leq K \), the quantity \( s_i \) is inductively given by

\[
s_i = \max \left\{ q : s_{i-1} + 1 \leq q \leq s \text{ and } |\bar{x}_q| > \frac{1}{\sqrt{2}} |\bar{x}_{s_{i-1}+1}| \right\}.
\]

In this way, we note that for any two index sets \( S_i \) and \( S_j \), \( S_i \cap S_j = \emptyset \) if \( i \neq j \). We also know by the definition of \( s_i \) that

\[
|\bar{x}_{s_{i+1}}| \leq \frac{1}{\sqrt{2}} |\bar{x}_{s_{i-1}+1}|, \; \forall 1 \leq i \leq K - 1. \tag{13}
\]

Now we show that after a finite number of iterations, say \( n \), the union of the \( S_i \)'s is contained in \( S^n \). To this end, we prove that for all \( 0 \leq i \leq K \),

\[
\bigcup_{i=0}^{i} S_i \subset S^{n_0 + n_1 + \cdots + n_i} \tag{14}
\]

for some \( n_i \)'s given below.

We pick \( n_0 = 0 \) and it is easy to verify that \( S_0 \subset S^0 \). Now suppose that (14) holds for \( i - 1 \). That is, the index set of the top \( s_{i-1} \) elements of \( \bar{x} \) is contained in \( S^{n_0 + \cdots + n_{i-1}} \). Due to Lemma 19, (14) holds for \( i \) as long as \( n_i \) satisfies

\[
\sqrt{2} |\bar{x}_{s_{i+1}}| > \alpha \gamma \cdot \beta^{n_i - 1} \|\bar{x}_{\{s_{i-1}+1, \ldots, s_i\}}\| + \theta. \tag{15}
\]

Note that

\[
\|\bar{x}_{\{s_{i-1}+1, \ldots, s_i\}}\|^2 = \|\bar{x}_{S_i}\|^2 + \cdots + \|\bar{x}_{S_K}\|^2 \\
\leq (\bar{x}_{s_{i-1}+1})^2 |S_i| + \cdots + (\bar{x}_{s_{i-1}+1})^2 \left( |S_K| + 2^{-1} |S_{i+1}| + \cdots + 2^{i-K} |S_K| \right) \\
< 2(\bar{x}_{s_i})^2 \left( |S_i| + 2^{-1} |S_{i+1}| + \cdots + 2^{i-K} |S_K| \right),
\]

where the second inequality follows from (13) and the last inequality follows from the definition of \( q_i \). Denote for simplicity

\[
T_i := |S_i| + 2^{-1} |S_{i+1}| + \cdots + 2^{i-K} |S_K|.
\]

As we assume \( \theta \leq \sqrt{2} \lambda |\bar{x}_{\text{min}}| \), we get

\[
\alpha \gamma \cdot \beta^{n_i - 1} \|\bar{x}_{\{s_{i-1}+1, \ldots, s_i\}}\| + \theta < \sqrt{2} \alpha \gamma |\bar{x}_{s_i}| \beta^{n_i - 1} \sqrt{T_i} + \sqrt{2} \lambda |\bar{x}_{s_i}|.
\]

Picking

\[
n_i = \log_{1/\beta} \frac{\alpha \gamma \sqrt{T_i}}{1 - \lambda} + 2
\]
Support Recovery of Hard Thresholding Pursuit

guarantees (15). It remains to calculate the total number of iterations. In fact, we have

\[ n = n_0 + n_1 + \ldots + n_K \]

\[ = \frac{1}{2 \log(1/\beta)} \sum_{i=1}^{K} \log T_i + K \cdot \frac{\log(\alpha \gamma/(1-\lambda))}{\log(1/\beta)} + 2K \]

\[ \zeta_1 \leq \frac{K}{2 \log(1/\beta)} \log \left( \frac{1}{K} \sum_{i=1}^{K} T_i \right) + \left( \frac{\log(\alpha \gamma/(1-\lambda))}{\log(1/\beta)} + 2 \right) K \]

\[ \zeta_2 \leq \frac{K}{2 \log(1/\beta)} \log \left( \frac{2}{K} \sum_{i=1}^{K} |S_i| \right) + \left( \frac{\log(\alpha \gamma/(1-\lambda))}{\log(1/\beta)} + 2 \right) K \]

\[ \zeta_3 \leq \left( \frac{\log 2}{2 \log(1/\beta)} + \frac{\log(\alpha \gamma/(1-\lambda))}{\log(1/\beta)} + 2 \right) s. \]

Above, \( \zeta_1 \) immediately follows by observing that the logarithmic function is concave. \( \zeta_2 \) uses the fact that after rearrangement, the coefficient of \( |S_i| \) is \( \sum_{j=0}^{i-1} 2^{-j} \) which is always smaller than 2. Finally, since the function \( r \log(2s/r) \) is monotonically increasing with respect to \( r \) and \( 1 \leq r \leq s \), \( \zeta_3 \) follows.

Combining this theorem, Lemma 19 and specific results in Prop. 2, Prop. 4 and Prop. 5 gives the main theorems in Section 3.