

## Abstract

This supplementary document contains the technical proofs of all the theorems in the NIPS'16 paper entitled "Exact Recovery of Hard Thresholding Pursuit". It is indeed the appendix section of the paper. The key technical lemmas are gathered in Appendix A, followed by the proofs of main results in Appendices B~ E.

## A Technical Lemmas

We present here a few technical lemmas to be used in our analysis.

**Lemma 1.** *Let  $x$  be a  $k$ -sparse vector and  $y = x - \eta \nabla f(x)$ . If  $f$  is  $M_{2k}$ -smooth, then the following inequality holds:*

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

*Proof.* Since  $f$  is  $M_{2k}$ -smooth, it follows that

$$\begin{aligned} f(y_k) - f(x) &\leq \langle \nabla f(x), y_k - x \rangle + \frac{M_{2k}}{2} \|y_k - x\|^2 \\ &\stackrel{\xi_1}{\leq} -\frac{1}{2\eta} \|y_k - x\|^2 + \frac{M_{2k}}{2} \|y_k - x\|^2 \\ &= -\frac{1 - \eta M_{2k}}{2\eta} \|y_k - x\|^2, \end{aligned}$$

where  $\xi_1$  follows from the fact that  $y_k$  is the best  $k$ -support approximation to  $y$  such that

$$\|y_k - y\|^2 = \|y_k - x + \eta \nabla f(x)\|^2 \leq \|x - x + \eta \nabla f(x)\|^2 = \|\eta \nabla f(x)\|^2,$$

which implies  $2\eta \langle \nabla f(x), y_k - x \rangle \leq -\|y_k - x\|^2$ . □

**Lemma 2.** *Assume that  $f$  is  $m_s$ -strongly convex. For any  $\|x - x'\|_0 \leq s$  it holds that*

$$\|x - x'\| \leq \sqrt{\frac{2 \max\{f(x) - f(x'), 0\}}{m_s}} + \frac{2\|\nabla_{F \cup F'} f(x')\|}{m_s},$$

where  $F = \text{supp}(x)$  and  $F' = \text{supp}(x')$ .

*Proof.* Since  $f$  is  $m_s$ -strongly convex, we have

$$\begin{aligned} f(x) &\geq f(x') + \langle \nabla f(x'), x - x' \rangle + \frac{m_s}{2} \|x - x'\|^2 \\ &\geq f(x') - \|\nabla_{F \cup F'} f(x')\| \|x - x'\| + \frac{m_s}{2} \|x - x'\|^2, \end{aligned}$$

where the second inequality follows from Cauchy-Schwarz inequality. From this above inequality we can see that if  $f(x) \leq f(x')$ , then

$$\|x - x'\| \leq \frac{2\|\nabla_{F \cup F'} f(x')\|}{m_s}.$$

If otherwise  $f(x) > f(x')$ , then we have

$$\begin{aligned} \|x - x'\| &\leq \frac{\|\nabla_{F \cup F'} f(x')\| + \sqrt{\|\nabla_{F \cup F'} f(x')\|^2 + 2m_s(f(x) - f(x'))}}{m_s} \\ &\leq \frac{2\|\nabla_{F \cup F'} f(x')\| + \sqrt{2m_s(f(x) - f(x'))}}{m_s}. \end{aligned}$$

This proves the bound. □

**Lemma 3.** Assume that  $f$  is  $m_s$ -strongly convex and  $M_s$ -smooth. For any index set  $F$  with cardinality  $|F| \leq s$  and any  $x, y$  with  $\text{supp}(x) \cup \text{supp}(y) \subseteq F$ , if  $\eta \in (0, 2m_s/M_s^2)$ , then

$$\|x - y - \eta \nabla_F f(x) + \eta \nabla_F f(y)\| \leq \sqrt{1 - 2\eta m_s + \eta^2 M_s^2} \|x - y\|,$$

and  $\sqrt{1 - 2\eta m_s + \eta^2 M_s^2} < 1$ .

*Proof.* By adding two copies of the inequality (2) with  $x$  and  $y$  interchanged and using the Theorem 2.1.5 in (Nesterov, 2004), we know that

$$(x - y)^\top (\nabla f(x) - \nabla f(y)) \geq m_s \|x - y\|^2, \quad \|\nabla_F f(x) - \nabla_F f(y)\| \leq M_s \|x - y\|.$$

For any  $\eta > 0$  we have

$$\|x - y - \eta \nabla_F f(x) + \eta \nabla_F f(y)\|^2 \leq (1 - 2\eta m_s + \eta^2 M_s^2) \|x - y\|^2.$$

Obviously  $1 - 2\eta m_s + \eta^2 M_s^2 \geq 1 - m_s^2/M_s^2 \geq 0$ , and  $\eta < 2m_s/M_s^2$  implies  $\sqrt{1 - 2\eta m_s + \eta^2 M_s^2} < 1$ . This proves the lemma.  $\square$

**Lemma 4.** Assume that  $f$  is  $M_s$ -smooth and  $m_s$ -strongly convex. Let  $F$  and  $F'$  be two index sets with cardinality  $|F \cup F'| = s$ . Let  $x = \arg \min_{\text{supp}(y) \subseteq F} f(y)$  and  $\text{supp}(x') \subseteq F'$ . Then for any  $\eta \in (0, 2m_s/M_s^2)$ , the following two inequalities hold

$$\|(x - x')_F\| \leq \frac{\rho \|x'_{F' \setminus F}\|}{1 - \rho} + \frac{\eta \|\nabla_{F \cup F'} f(x')\|}{(1 - \rho)}, \quad (\text{A.1})$$

$$\|x - x'\| \leq \frac{\|x'_{F' \setminus F}\|}{1 - \rho} + \frac{\eta \|\nabla_{F \cup F'} f(x')\|}{(1 - \rho)}, \quad (\text{A.2})$$

where  $\rho = \sqrt{1 - 2\eta m_s + \eta^2 M_s^2} < 1$ .

*Proof.* Since  $x$  is the minimum of  $f(y)$  restricted over the supporting set  $F$ , we have  $\langle \nabla f(x), z \rangle = 0$  whenever  $\text{supp}(z) \subseteq F$ . It follows that

$$\begin{aligned} & \|(x - x')_F\|^2 \\ &= \langle x - x', (x - x')_F \rangle \\ &= \langle x - x' - \eta \nabla_{F \cup F'} f(x) + \eta \nabla_{F \cup F'} f(x'), (x - x')_F \rangle - \eta \langle \nabla_{F \cup F'} f(x'), (x - x')_F \rangle \\ &\stackrel{\xi_1}{\leq} \sqrt{1 - 2\eta m_s + \eta^2 M_s^2} \|x - x'\| \|(x - x')_F\| + \eta \|\nabla_{F \cup F'} f(x')\| \|(x - x')_F\|, \end{aligned}$$

where  $\xi_1$  uses Lemma 3. Let us abbreviate  $\rho = \sqrt{1 - 2\eta m_s + \eta^2 M_s^2}$ . After simplification, we have

$$\|(x - x')_F\| \leq \rho \|x - x'\| + \eta \|\nabla_{F \cup F'} f(x')\|. \quad (\text{A.3})$$

It follows that

$$\begin{aligned} \|x - x'\| &\leq \|(x - x')_F\| + \|(x - x')_{F' \setminus F}\| \\ &\leq \rho \|x - x'\| + \eta \|\nabla_{F \cup F'} f(x')\| + \|(x - x')_{F' \setminus F}\|. \end{aligned}$$

After rearrangement we obtain

$$\begin{aligned} \|x - x'\| &\leq \frac{\|(x - x')_{F' \setminus F}\|}{1 - \rho} + \frac{\eta \|\nabla_{F \cup F'} f(x')\|}{1 - \rho} \\ &= \frac{\|x'_{F' \setminus F}\|}{1 - \rho} + \frac{\eta \|\nabla_{F \cup F'} f(x')\|}{1 - \rho}. \end{aligned} \quad (\text{A.4})$$

By combining (A.3) and (A.4) we get

$$\|(x - x')_F\| \leq \frac{\rho \|x'_{F' \setminus F}\|}{1 - \rho} + \frac{\eta \|\nabla_{F \cup F'} f(x')\|}{1 - \rho}.$$

This proves the desired bounds in this lemma.  $\square$

## B Proof of Proposition 1

The following lemma gives a necessary condition on the value  $x_{\min}^*$ .

**Lemma 5.** *If  $f$  is  $M_{2k}$ -smooth, then the following inequality holds for the global minimizer  $x^*$ :*

$$x_{\min}^* \geq \frac{\|\nabla f(x^*)\|_{\infty}}{M_{2k}}.$$

*Proof.* Assume otherwise that  $\vartheta^* := \frac{M_{2k}x_{\min}^*}{\|\nabla f(x^*)\|_{\infty}} < 1$ . Let us consider  $\tilde{x}^* = x^* - \eta \nabla f(x^*)$  with any  $\eta \in (\vartheta^*/M_{2k}, 1/M_{2k})$ . From Lemma 1 we get that

$$f(\tilde{x}_k^*) \leq f(x^*) - \frac{1 - \eta M_{2k}}{2\eta} \|\tilde{x}_k^* - x^*\|^2.$$

Since  $\eta < \frac{1}{M_{2k}}$  and  $x_{\min}^* = \frac{\vartheta^* \|\nabla f(x^*)\|_{\infty}}{M_{2k}} < \eta \|\nabla f(x^*)\|_{\infty}$ , we have  $\tilde{x}_k^* \neq x^*$  and thus it follows from the above inequality that  $f(\tilde{x}_k^*) < f(x^*)$  which contradicts the optimality of  $x^*$ .  $\square$

We also need the following lemma in our proof.

**Lemma 6.** *Assume that  $f$  is  $M_{2k}$ -smooth and  $m_{2k}$ -strongly convex. Let  $\bar{x}$  be an arbitrary  $k$ -sparse vector satisfying  $\bar{x} = \arg \min_{\text{supp}(x) \subseteq \text{supp}(\bar{x})} f(x)$  and  $\bar{x}_{\min} \geq \frac{\bar{\vartheta}}{M_{2k}} \|\nabla f(\bar{x})\|_{\infty}$  for some  $\bar{\vartheta} > 1$ . If  $\frac{m_{2k}}{M_{2k}} \geq \max\left\{\frac{3\bar{\vartheta}+1}{4\bar{\vartheta}}, \frac{\sqrt{3}}{2}\right\}$ , then  $\bar{x}$  is the global minimizer, i.e.,  $\bar{x} = x^*$ .*

*Proof.* Denote  $\bar{F} = \text{supp}(\bar{x})$ . Assume otherwise that  $\bar{F} \neq F^*$ . By using the inequality (A.1) in Lemma 4 and the definition of  $\bar{\vartheta}$  we have

$$\begin{aligned} \|x_{F^* \setminus \bar{F}}^*\| &\leq \|(x^* - \bar{x})_{F^*}\| \\ &\leq \frac{\rho \|\bar{x}_{\bar{F} \setminus F^*}\|}{1 - \rho} + \frac{m_{2k} \|\nabla_{F^* \setminus \bar{F}} f(\bar{x})\|}{(1 - \rho) M_{2k}^2} \\ &\leq \frac{(\bar{\vartheta} \rho + m_{2k}/M_{2k}) \|\bar{x}_{\bar{F} \setminus F^*}\|}{\bar{\vartheta}(1 - \rho)} \\ &\stackrel{\xi_1}{\leq} \left(1 + \frac{2m_{2k}}{\bar{\vartheta} M_{2k}}\right) \|\bar{x}_{\bar{F} \setminus F^*}\|, \end{aligned} \tag{A.5}$$

where the last inequality  $\xi_1$  follows from  $\rho \leq 0.5$  (as  $m_{2k}/M_{2k} \geq \frac{\sqrt{3}}{2}$ ). Let  $F = F^* \cup \bar{F}$ . Since  $f$  is  $m_{2k}$ -strongly convex, it can be easily verified that

$$\|\nabla_F f(\bar{x}) - \nabla_F f(x^*)\| \geq m_{2k} \|\bar{x} - x^*\|. \tag{A.6}$$

It follows that

$$\begin{aligned} &\|\nabla_{F^* \setminus \bar{F}} f(\bar{x})\| + \|\nabla_{\bar{F} \setminus F^*} f(x^*)\| \\ &\stackrel{\xi_1}{\leq} \|\nabla_F f(\bar{x})\| + \|\nabla_F f(x^*)\| \\ &\stackrel{\xi_2}{\geq} \|\nabla_F f(\bar{x}) - \nabla_F f(x^*)\| \\ &\stackrel{\xi_3}{\geq} m_{2k} \|\bar{x} - x^*\| \\ &\geq m_{2k} \left( \|\bar{x}_{\bar{F} \setminus F^*}\| + \|x_{F^* \setminus \bar{F}}^*\| \right), \end{aligned} \tag{A.7}$$

where the inequality  $\xi_1$  follows from the optimality condition, i.e.,  $\nabla_{F^*} f(x^*) = \nabla_{\bar{F}} f(\bar{x}) = 0$ ; the inequality  $\xi_2$  follows from triangle inequality; the inequality  $\xi_3$  follows from (A.6). Since the

definition of  $\bar{\vartheta}$  implies  $\|\nabla_{F^* \setminus \bar{F}} f(\bar{x})\| \leq \frac{M_{2k}}{\bar{\vartheta}} \|\bar{x}_{\bar{F} \setminus F^*}\|$ , it follows from the inequality (A.7) that

$$\begin{aligned} \|\nabla_{\bar{F} \setminus F^*} f(x^*)\| &\geq \left( m_{2k} - \frac{M_{2k}}{\bar{\vartheta}} \right) \|\bar{x}_{\bar{F} \setminus F^*}\| + m_{2k} \|x_{F^* \setminus \bar{F}}^*\| \\ &\stackrel{\xi_1}{\geq} \left( \frac{m_{2k} - M_{2k}/\bar{\vartheta}}{1 + 2m_{2k}/(\bar{\vartheta}M_{2k})} + m_{2k} \right) \|x_{F^* \setminus \bar{F}}^*\| \\ &\stackrel{\xi_2}{>} \left( \frac{m_{2k} - M_{2k}/\bar{\vartheta}}{3} + m_{2k} \right) \|x_{F^* \setminus \bar{F}}^*\| \\ &= \left( \frac{4m_{2k} - M_{2k}/\bar{\vartheta}}{3} \right) \|x_{F^* \setminus \bar{F}}^*\| \\ &\stackrel{\xi_3}{\geq} M_{2k} \|x_{F^* \setminus \bar{F}}^*\|, \end{aligned}$$

where inequality  $\xi_1$  follows from (A.5) and the assumption on  $m_{2k}/M_{2k}$  which implies  $m_{2k} > \frac{M_{2k}}{\bar{\vartheta}}$ ,  $\xi_2$  follows from  $\bar{\vartheta} > 1$  and  $m_{2k} < M_{2k}$ , and  $\xi_3$  follows from  $m_{2k}/M_{2k} \geq \frac{3\bar{\vartheta}+1}{4\bar{\vartheta}}$ . The previous inequality obviously leads to

$$x_{\min}^* < \frac{\|\nabla f(x^*)\|_{\infty}}{M_{2k}}.$$

This contradicts the x-min condition we have proved for  $x^*$  in Lemma 5. Therefore, we must have  $\bar{x} = x^*$  holds.  $\square$

Now we can prove Proposition 1.

*Proof of Proposition 1.* We first show that  $\text{supp}(\bar{x}) = \text{supp}(x^*)$  if Condition (1) holds. Assume otherwise  $\text{supp}(\bar{x}) \neq \text{supp}(x^*)$ . From the optimality of  $x^*$  we have  $f(x^*) \leq f(\bar{x})$ . By invoking Lemma 2 we get

$$\bar{x}_{\min} < \|x^* - \bar{x}\| \leq \frac{2\sqrt{2k}\|\nabla f(\bar{x})\|_{\infty}}{m_{2k}},$$

which contradicts the condition.

Next we show that  $\text{supp}(\bar{x}) = \text{supp}(x^*)$  if Condition (2) is satisfied. From the definition of  $\bar{x}^*$  and  $\bar{l}$  we have

$$\begin{aligned} f(\bar{x}) &\geq f(\bar{x}^*) \\ &\geq f(\bar{x}) - \|\nabla f(\bar{x})\|_{\infty} \|\bar{x}^* - \bar{x}\|_1 + \frac{\bar{l}}{2} \|\bar{x}^* - \bar{x}\|_1^2 \\ &\geq f(\bar{x}) - \frac{\|\nabla f(\bar{x})\|_{\infty}^2}{\bar{l}} - \frac{\bar{l}}{4} \|\bar{x}^* - \bar{x}\|_1^2 + \frac{\bar{l}}{2} \|\bar{x}^* - \bar{x}\|_1^2 \\ &= f(\bar{x}) - \frac{\|\nabla f(\bar{x})\|_{\infty}^2}{\bar{l}} + \frac{\bar{l}}{4} \|\bar{x}^* - \bar{x}\|_1^2. \end{aligned}$$

Therefore,

$$\|\bar{x}^* - \bar{x}\|_1 \leq \frac{2\|\nabla f(\bar{x})\|_{\infty}}{\bar{l}}. \quad (\text{A.8})$$

Let  $\bar{F} = \text{supp}(\bar{x})$ . Since  $f$  is  $M_{2k}$ -smooth, we know from (Nesterov, 2004, Theorem 2.1.5) and the above inequality that

$$\|\nabla f(\bar{x}^*) - \nabla f(\bar{x})\| \leq M_{2k} \|\bar{x}^* - \bar{x}\| \leq M_{2k} \|\bar{x}^* - \bar{x}\|_1 \leq \frac{2M_{2k}\|\nabla f(\bar{x})\|_{\infty}}{\bar{l}}, \quad (\text{A.9})$$

where we have also used the fact  $\|x\| \leq \|x\|_1$ . By applying triangle inequality to (A.8) and (A.9) we further get

$$\bar{x}_{\min}^* \geq \bar{x}_{\min} - \frac{2\|\nabla f(\bar{x})\|_{\infty}}{\bar{l}}, \quad \|\nabla f(\bar{x}^*)\|_{\infty} \leq \frac{(\bar{l} + 2M_{2k})\|\nabla f(\bar{x})\|_{\infty}}{\bar{l}}.$$

Given the condition on  $\bar{x}_{\min}$ , we can obtain from the above

$$\bar{x}_{\min}^* \geq \frac{\bar{\vartheta}\|\nabla f(\bar{x}^*)\|_{\infty}}{M_{2k}}.$$

Thus, by using Lemma 6 we obtain that  $\bar{x}^* = x^*$ . This leads the desired result since  $\text{supp}(\bar{x}^*) = \text{supp}(\bar{x})$ .  $\square$

## C Proof of Theorem 1

In large picture, the proof of Theorem 1 contains the following three key steps:

- (a) In the first step, we prove that under the conditions stated in the theorem, HTP terminates when  $x^{(t)}$  reaches  $x^*$ .
- (b) In the second step, we show that if  $x^{(t-1)} \neq x^*$ , then HTP will output  $x^{(t)} \neq x^{(t-1)}$ . That is, the algorithm will not terminate before reaching  $x^*$ .
- (c) In the final step, we show that the conditions in the theorem guarantee a unique optimal solution  $x^*$  and finite termination of HTP.

*Proof of Theorem 1. Step (a):* We first prove that under the conditions stated in the theorem, the algorithm will terminate when it reaches  $x^*$ . Indeed, let us assume  $x^{(t-1)} = x^*$ . Since  $\vartheta^* > 1$  and  $\eta = m_{2k}/M_{2k}^2 < 1/M_{2k}$ , we obtain that

$$x_{\min}^{(t-1)} = x_{\min}^* = \frac{\theta^* \|\nabla f(x^*)\|_{\infty}}{M_{2k}} > \frac{\|\nabla f(x^*)\|_{\infty}}{M_{2k}} > \eta \|\nabla f(x^*)\|_{\infty} = \eta \nabla_{\max}^{(t-1)},$$

which indicates that the top  $k$  (in magnitude) entries of  $\tilde{x}^{(t)} = x^{(t-1)} - \eta \nabla f(x^{(t-1)})$  are exactly the  $k$  nonzero entries of  $x^{(t-1)}$  and thus  $F^{(t)} = F^{(t-1)}$ . The step **S3** of HTP ensures  $x^{(t)} = x^{(t-1)}$ . Therefore, HTP terminates at  $x^{(t-1)} = x^*$ .

**Step (b):** Further, we show that at time instance  $t - 1$ , if  $x^{(t-1)} \neq x^*$ , then HTP will output  $x^{(t)} \neq x^{(t-1)}$ . From the inequality (A.1) and the definition of  $\vartheta^*$  we have

$$\begin{aligned} \|x_{F^{(t-1)} \setminus F^*}^{(t-1)}\| &\leq \| (x^{(t-1)} - x^*)_{F^{(t-1)}} \| \\ &\leq \frac{\rho \| (x_{F^* \setminus F^{(t-1)}}^* \|}{1 - \rho} + \frac{\eta \|\nabla_{F^{(t-1)} \setminus F^*} f(x^*)\|}{1 - \rho} \\ &\leq \frac{(\rho \vartheta^* + \eta M_{2k}) \|x_{F^* \setminus F^{(t-1)}}^*\|}{\vartheta^* (1 - \rho)} \\ &\stackrel{\xi_1}{\leq} \left(1 + \frac{2m_{2k}}{\vartheta^* M_{2k}}\right) \|x_{F^* \setminus F^{(t-1)}}^*\|, \end{aligned} \tag{A.10}$$

where the last inequality  $\xi_1$  follows from  $\eta = m_{2k}/M_{2k}^2$  and  $\rho \leq 0.5$  (as  $m_{2k}/M_{2k} \geq \frac{7\vartheta^* + 1}{8\vartheta^*} > \frac{\sqrt{3}}{2}$ ). Let  $F = F^* \cup F^{(t-1)}$ . Since  $f$  is  $m_{2k}$ -strongly convex, it can be easily verified that

$$\|\nabla_F f(x^*) - \nabla_F f(x^{(t-1)})\| \geq m_{2k} \|x^* - x^{(t-1)}\|. \tag{A.11}$$

Thus

$$\begin{aligned} &\|\nabla_{F^{(t-1)} \setminus F^*} f(x^*)\| + \|\nabla_{F^* \setminus F^{(t-1)}} f(x^{(t-1)})\| \\ &\stackrel{\xi_1}{=} \|\nabla_F f(x^*)\| + \|\nabla_F f(x^{(t-1)})\| \\ &\stackrel{\xi_2}{\geq} \|\nabla_F f(x^*) - \nabla_F f(x^{(t-1)})\| \\ &\stackrel{\xi_3}{\geq} m_{2k} \|x^* - x^{(t-1)}\| \\ &\geq m_{2k} \left( \|x_{F^* \setminus F^{(t-1)}}^*\| + \|x_{F^{(t-1)} \setminus F^*}^{(t-1)}\| \right), \end{aligned} \tag{A.12}$$

where the inequality  $\xi_1$  follows from the optimality condition of  $\nabla_{F^*} f(x^*) = \nabla_{F^{(t-1)}} f(x^{(t-1)}) = 0$ ; the inequality  $\xi_2$  follows from triangle inequality; the inequality  $\xi_3$  follows from (A.11). Since the definition of  $\vartheta^*$  implies  $\|\nabla_{F^{(t-1)} \setminus F^*} f(x^*)\| \leq \frac{M_{2k}}{\vartheta^*} \|x_{F^* \setminus F^{(t-1)}}^*\|$ , it follows from the inequal-

ity (A.12) that

$$\begin{aligned}
\|\nabla_{F^* \setminus F^{(t-1)}} f(x^{(t-1)})\| &\geq \left(m_{2k} - \frac{M_{2k}}{\vartheta^*}\right) \|x_{F^* \setminus F^{(t-1)}}^*\| + m_{2k} \|x_{F^{(t-1)} \setminus F^*}^{(t-1)}\| \\
&\stackrel{\xi_1}{\geq} \left(\frac{m_{2k} - M_{2k}/\vartheta^*}{1 + 2m_{2k}/(\vartheta^* M_{2k})} + m_{2k}\right) \|x_{F^{(t-1)} \setminus F^*}^{(t-1)}\| \\
&\stackrel{\xi_2}{>} \left(\frac{m_{2k} - M_{2k}/\vartheta^*}{3} + m_{2k}\right) \|x_{F^{(t-1)} \setminus F^*}^{(t-1)}\| \\
&= \left(\frac{4m_{2k} - M_{2k}/\vartheta^*}{3}\right) \|x_{F^{(t-1)} \setminus F^*}^{(t-1)}\|,
\end{aligned}$$

where inequality  $\xi_1$  follows from (A.10) and the assumption on  $m_{2k}/M_{2k}$  which implies  $m_{2k} > \frac{M_{2k}}{\vartheta^*}$ , and  $\xi_2$  follows from  $\vartheta^* > 1$  and  $m_{2k} < M_{2k}$ . Now we claim that  $\eta \|\nabla_{F^* \setminus F^{(t-1)}} f(x^{(t-1)})\| > \|x_{F^{(t-1)} \setminus F^*}^{(t-1)}\|$ . Indeed, since  $\eta = \frac{m_{2k}}{M_{2k}^2}$  and  $\frac{m_{2k}}{M_{2k}} \geq \frac{1+7\vartheta^*}{8\vartheta^*}$ , it can be verified from the previous inequality that

$$\eta \|\nabla_{F^* \setminus F^{(t-1)}} f(x^{(t-1)})\| > \left(\frac{4m_{2k}^2}{3M_{2k}^2} - \frac{m_{2k}}{3M_{2k}\vartheta^*}\right) \|x_{F^{(t-1)} \setminus F^*}^{(t-1)}\| \geq \|x_{F^{(t-1)} \setminus F^*}^{(t-1)}\|.$$

Thus at least the nonzero entries of  $x^{(t-1)}$  supported on  $F^{(t-1)} \setminus F^*$  and the entries of  $\nabla f(x^{(t-1)})$  supported on  $F^* \setminus F^{(t-1)}$  can (but not have to) be swapped in the step **S2**. Thus  $F^{(t)} \neq F^{(t-1)}$  and HTP proceeds to  $x^{(t)} \neq x^{(t-1)}$ .

**Step (c):** The final step is to further show that under the conditions stated in the theorem, the optimal solution  $x^*$  is unique and the sequence  $\{f(x^{(t)})\}$  generated by HTP reach  $f(x^*)$  within a finite number of iteration. We first prove the uniqueness of optimal solution by contradiction. Assume otherwise there exists  $x^{**} \neq x^*$  such that  $f(x^{**}) = f(x^*)$ . Let  $F^{**} = \text{supp}(x^{**})$ . Since  $f$  is  $m_{2k}$ -strongly convex, it is true that  $F^{**} \neq F^*$ . Similar to the arguments in part(b) we can prove that

$$\eta \|\nabla_{F^* \setminus F^{**}} f(x^{**})\| > \|x_{F^{**} \setminus F^*}^{**}\|.$$

Now consider  $\tilde{x}^{**} = x^{**} - \eta \nabla f(x^{**})$ . From the above inequality we get  $\tilde{x}_k^{**} \neq x_k^{**}$ . By using Lemma 1 we get

$$f(\tilde{x}_k^{**}) - f(x^*) = f(\tilde{x}_k^{**}) - f(x^{**}) \leq -\frac{1 - \eta M_{2k}}{2\eta} \|\tilde{x}_k^{**} - x_k^{**}\|^2 < 0.$$

This contradicts the optimality of  $x^*$ . Therefore, the optimal solution  $x^*$  is unique under the conditions in the theorem.

Last, we show that  $f(x^{(t)})$  will reach  $f(x^*)$  in finite steps. In the above Step (b) we have proved  $x^{(t)} \neq x^{(t-1)}$  whenever  $x^{(t)} \neq x^*$ , and in such a case it must hold that  $|F^* \setminus F^{(t-1)}| \leq |F^{(t)} \setminus F^{(t-1)}|$ . Thus from the fact that  $\nabla_{F^{(t)} \setminus F^{(t-1)}} f(x^{(t-1)})$  contains the top  $|F^{(t)} \setminus F^{(t-1)}|$  entries of  $\nabla f(x^{(t-1)})$  we get

$$\|\nabla_{F^* \setminus F^{(t-1)}} f(x^{(t-1)})\| \leq \|\nabla_{F^{(t)} \setminus F^{(t-1)}} f(x^{(t-1)})\|.$$

From the definition of  $\tilde{x}^{(t)}$  we have that the following inequality holds:

$$\|\tilde{x}_{F^{(t)}}^{(t)} - x^{(t-1)}\| \geq \eta \|\nabla_{F^{(t)} \setminus F^{(t-1)}} f(x^{(t-1)})\|. \quad (\text{A.13})$$

By using Lemma 1, (A.13) and the above inequality we have

$$f(x^{(t)}) - f(x^{(t-1)}) \leq f(\tilde{x}_{F^{(t)}}^{(t)}) - f(x^{(t-1)}) \leq -\frac{(1 - \eta M_{2k})\eta}{2} \|\nabla_{F^* \setminus F^{(t-1)}} f(x^{(t-1)})\|^2. \quad (\text{A.14})$$

From the  $m_{2k}$ -strong convexity of  $f$  we have

$$\begin{aligned}
\frac{m_{2k}}{2} \|x^* - x^{(t-1)}\|^2 &\leq f(x^*) - f(x^{(t-1)}) - (x^* - x^{(t-1)})^\top \nabla f(x^{(t-1)}) \\
&\stackrel{\xi_1}{\leq} f(x^*) - f(x^{(t-1)}) + \frac{m_{2k}}{2} \|x^* - x^{(t-1)}\|^2 + \frac{1}{2m_{2k}} \|\nabla_{F^* \setminus F^{(t-1)}} f(x^{(t-1)})\|^2,
\end{aligned}$$

where  $\xi_1$  follows from Cauchy-Schwartz inequality,  $ma^2/2 + b^2/(2m) \geq ab$  for any  $m > 0$ , and  $\nabla_{F^{(t-1)}} f(x^{(t-1)}) = 0$ . This implies

$$\|\nabla_{F^* \setminus F^{(t-1)}} f(x^{(t-1)})\|^2 \geq 2m_{2k} [f(x^{(t-1)}) - f(x^*)]. \quad (\text{A.15})$$

By combining (A.14) and (A.15) we arrive at

$$f(x^{(t)}) - f(x^{(t-1)}) \leq -(1 - \eta M_{2k}) m_{2k} \eta (f(x^{(t-1)}) - f(x^*)) = -\left(1 - \frac{m_{2k}}{M_{2k}}\right) \frac{m_{2k}^2}{M_{2k}^2} (f(x^{(t-1)}) - f(x^*)).$$

Therefore, we get

$$f(x^{(t)}) - f(x^*) \leq (1 - \nu)(f(x^{(t-1)}) - f(x^*)),$$

where

$$\nu = \left(1 - \frac{m_{2k}}{M_{2k}}\right) \frac{m_{2k}^2}{M_{2k}^2}.$$

Therefore  $f(x^{(t)}) - f(x^*) \leq \Delta^{-*}$  when  $t \geq \frac{1}{\nu} \ln \frac{\Delta^{(0)}}{\Delta^{-*}}$  (note that  $\Delta^{-*} > 0$  due to the uniqueness of optimal solution). After that, we have  $f(x^{(t)}) < f(x^*)$  and thus  $f(x^{(t)}) = f(x^*)$ . Since  $x^*$  is unique, it must hold that  $x^{(t)} = x^*$ .

By combining the above three steps (a), (b) and (c), we complete the proof of Theorem 1.  $\square$

## D Proof of Theorem 2

*Proof.* Since the conditions of Theorem 1 are assumed to be fulfilled, it follows from the theorem that  $x^{(t)} = x^*$  after sufficient iteration. If  $\bar{x}_{\min} \geq \frac{2\sqrt{2k}}{m_{2k}} \|\nabla f(\bar{x})\|_{\infty}$ , then we know from Proposition 1 that  $\text{supp}(\bar{x}) = \text{supp}(x^*)$ , and thus  $\text{supp}(x^{(t)}) = \text{supp}(\bar{x})$ . If  $\bar{x}_{\min} \geq \left(\frac{\vartheta^*}{M_{2k}} + \frac{2\vartheta^*+2}{l}\right) \|\nabla f(\bar{x})\|_{\infty}$ , then using the arguments in the proof of Proposition 1 we can derive that

$$\bar{x}_{\min}^* \geq \frac{\vartheta^* \|\nabla f(\bar{x}^*)\|_{\infty}}{M_{2k}}.$$

Since  $\frac{m_{2k}}{M_{2k}} \geq \frac{7\vartheta^*+1}{8\vartheta^*}$ , we get from Lemma 6 that  $\bar{x}^* = x^*$ . This leads to the desired result as  $\text{supp}(x^{(t)}) = \text{supp}(x^*) = \text{supp}(\bar{x}^*) = \text{supp}(\bar{x})$ .  $\square$

## E Proof of Theorem 3

*Proof of Theorem 3. Part(a).* This part can be proved by showing that  $F^{(t)} \neq F^{(t-1)}$  whenever  $\text{supp}(\bar{x}) \not\subseteq \text{supp}(x^{(t-1)})$ . To this end, let us assume  $\text{supp}(\bar{x}) \not\subseteq \text{supp}(x^{(t-1)})$ . Then

$$\begin{aligned} \bar{x}_{\min} + \|x_{F^{(t-1)} \setminus \bar{F}}^{(t-1)}\| &\leq \|\bar{x} - x^{(t-1)}\| \\ &\leq \xi_1 \sqrt{\frac{2 \max\{f(\bar{x}) - f(x^{(t-1)}), 0\}}{m_{2k}}} + \frac{2\|\nabla_{\bar{F} \setminus F^{(t-1)}} f(x^{(t-1)})\|}{m_{2k}} \\ &\leq \xi_2 \sqrt{\frac{2(f(\bar{x}) - f(x^*))}{m_{2k}}} + \frac{2\|\nabla_{\bar{F} \setminus F^{(t-1)}} f(x^{(t-1)})\|}{m_{2k}}, \end{aligned}$$

where “ $\xi_1$ ” follows from Lemma 2 and “ $\xi_2$ ” is due to the fact of  $f(x^{(t-1)}) \geq f(x^*)$ . Since it is assumed  $\bar{x}_{\min} > \sqrt{\frac{2(f(\bar{x}) - f(x^*))}{m_{2k}}}$ , the above inequality implies that

$$\|x_{F^{(t-1)} \setminus \bar{F}}^{(t-1)}\| < \frac{2\|\nabla_{\bar{F} \setminus F^{(t-1)}} f(x^{(t-1)})\|}{m_{2k}}.$$

Let  $x_{\text{bot}(|\bar{F} \setminus F^{(t-1)}|)}^{(t-1)}$  be the smallest (in magnitude)  $|\bar{F} \setminus F^{(t-1)}|$  elements of  $x^{(t-1)}$ . Then it can be shown from the previous inequality that

$$\|\nabla_{\bar{F} \setminus F^{(t-1)}} f(x^{(t-1)})\| > \frac{m_{2k} |\bar{F}^{(t-1)} \setminus \bar{F}|^{1/2}}{2|\bar{F} \setminus F^{(t-1)}|^{1/2}} \|x_{\text{bot}(|\bar{F} \setminus F^{(t-1)}|)}^{(t-1)}\| \geq \frac{(k - \bar{k})^{1/2} m_{2k}}{2\bar{k}^{1/2}} \|x_{\text{bot}(|\bar{F} \setminus F^{(t-1)}|)}^{(t-1)}\|.$$

Since  $\eta = \frac{1}{2M_{2k}}$  and  $k \geq \left(1 + \frac{16M_{2k}^2}{m_{2k}^2}\right) \bar{k}$ , we thus have

$$\eta \|\nabla_{\bar{F} \setminus F^{(t-1)}} f(x^{(t-1)})\| > \|x_{\text{bot}(|\bar{F} \setminus F^{(t-1)}|)}^{(t-1)}\|.$$

This further implies at least the bottom  $|\bar{F} \setminus F^{(t-1)}|$  entries of  $x^{(t-1)}$  can (but not have to) be swapped in the step **S2** of Algorithm 1, and thus  $F^{(t)} \neq F^{(t-1)}$ . Therefore, when the algorithm terminates at time instance  $t$ , i.e.,  $F^{(t+1)} = F^{(t)}$ , we must have  $\text{supp}(\bar{x}) \subseteq \text{supp}(x^{(t)})$  holds.

Finally, we claim that HTP is finite under the assumed conditions. Indeed, based on Lemma 1 it is easy to verify that when  $\eta = \frac{1}{M_{2k}}$ , the sequence  $\{f(x^{(t)})\}$  generated by Algorithm 1 is monotone. Since the number of  $k$ -support index sets is finite, the sequence  $\{f(x^{(t)})\}$  will be eventually periodic, and thus must be eventually a constant. Therefore we deduce that  $\tilde{x}_k^{(t)} = x^{(t-1)}$ , i.e.,  $F^{(t)} = F^{(t-1)}$ , for  $t$  large enough. This completes the proof of Part (a).

**Part(b):** Given the conditions in this part, we know from Part(a) that  $\text{supp}(\bar{x}) \subseteq \text{supp}(x^{(t)})$  when HTP terminates at time instance  $t$ . Let us assume  $\text{supp}(\bar{x}) \neq \text{supp}(x_k^{(t)})$ . Then

$$\begin{aligned} \bar{x}_{\min} &\leq \|\bar{x} - x_k^{(t)}\| \\ &\stackrel{\xi_1}{\leq} 1.62 \|\bar{x} - x^{(t)}\| \\ &\stackrel{\xi_2}{\leq} 1.62 \sqrt{\frac{2(f(\bar{x}) - f(x^*))}{m_{2k}}} + \frac{2\|\nabla_{\bar{F} \setminus F^{(t)}} f(x^{(t)})\|}{m_{2k}} \\ &\stackrel{\xi_3}{\leq} 1.62 \sqrt{\frac{2(f(\bar{x}) - f(x^*))}{m_{2k}}}, \end{aligned}$$

where “ $\xi_1$ ” is based on the truncation error bound in (Shen & Li, 2016, Theorem 1), “ $\xi_2$ ” follows from Lemma 2 and the fact of  $f(x^{(t)}) \geq f(x^*)$ , and “ $\xi_3$ ” is the consequence of  $\bar{F} \subseteq F^{(t)}$  and the optimality of  $x^{(t)}$  over  $F^{(t)}$ . This above inequality contradicts the assumption on  $\bar{x}_{\min}$ . Therefore, it must holds that  $\text{supp}(\bar{x}) = \text{supp}(x_k^{(t)})$ .  $\square$