A   Proofs related to OMPR: Exact Recovery Case

Let us denote the objective function by $f(x) = \frac{1}{2}\|Ax - b\|^2$. Let I_t denote the support set of x^t and $I^\star$ be the support set of $x^\star$. Define the sets

$$
\begin{aligned}
FA_t &= I_t \backslash I^\star & &\text{(false alarms)} \\
MD_t &= I^\star \backslash I_t & &\text{(missed detections)} \\
CO_t &= I_t \cap I^\star & &\text{(correct detections)} .
\end{aligned}
$$

As the algorithms proceed, elements move in and out of the current set I_t. Let us give names to the set of found and lost elements as we move from I_t to I_{t+1}:

$$
\begin{aligned}
F_t &= I_{t+1} \backslash I_t & &\text{(found)} \\
L_t &= I_t \backslash I_{t+1} & &\text{(lost) .}
\end{aligned}
$$

We first state two technical lemmas that we will need. These can be found in [19].

Lemma 13. *For any $S \subset [n]$, we have,*

$$
\|I - A_S^T A_S\| \le \delta_{|S|}.
$$

Lemma 14. *For any $S, T \subset [n]$ such that $S \cap T = \emptyset$, we have,*

$$
\|A_S^T A_T\|_2 \le \delta_{|S \cup T|}.
$$

Proof of Theorem 4

Lemma 15. *Let $\delta_{2k} < 1 - \frac{1}{2\eta}$, $f(x^t) > 0$. Then, in OMPR (l),*

$$
0 < 2(2\eta - \frac{1}{1 - \delta_{2k}})f(x^t) \le \|z_{MD_t}^{t+1}\|^2 - \|x_{FA_t}^t\|^2.
$$

Proof. Since $x_{I_t}^t$ is the solution to the least squares problem $\min_x \|A_{I_t} x - b\|^2$,

$$
A_{I_t}^T (A_{I_t} x_{I_t}^t - b) = \mathbf{0}. \tag{6}
$$

Now, note that

$$
\begin{aligned}
f(x^t) &= \frac{1}{2}\|A_{I_t} x_{I_t}^t - A_{I^\star} x_{I^\star}^\star\|^2, \\
&= \frac{1}{2}((x_{I_t}^t)^T A_{I_t}^T (A_{I_t} x_{I_t}^t - A_{I^\star} x_{I^\star}^\star) - (x_{I^\star}^\star)^T A_{I^\star}^T (A_{I_t} x_{I_t}^t - A_{I^\star} x_{I^\star}^\star)), \\
&= -\frac{1}{2}(x_{MD_t}^\star)^T A_{MD_t}^T (A_{I_t} x_{I_t}^t - A_{I^\star} x_{I^\star}^\star), &&\text{by (6)} \\
&= \frac{1}{2\eta}(x_{MD_t}^\star)^T z_{MD_t}^{t+1}. &&\text{by (3)} \tag{7}
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\|x_{MD_t}^\star - z_{MD_t}^{t+1}\|^2 &= \|x_{MD_t}^\star\|^2 + \|z_{MD_t}^{t+1}\|^2 - 2(x_{MD_t}^\star)^T z_{MD_t}^{t+1} \\
&= \|x_{MD_t}^\star\|^2 + \|z_{MD_t}^{t+1}\|^2 - 4\eta f(x^t). \tag{8}
\end{aligned}
$$

That is,

$$
\begin{aligned}
4\eta f(x^t) &\le \|x_{MD_t}^\star\|^2 + \|z_{MD_t}^{t+1}\|^2, \\
&\le \|x_{MD_t}^\star\|^2 + \|x_{FA_t}^t\|^2 + \|x_{CO_t}^t - x_{CO_t}^\star\|^2 - \|x_{FA_t}^t\|^2 + \|z_{MD_t}^{t+1}\|^2, \\
&= \|x^t - x^\star\|^2 + \|z_{MD_t}^{t+1}\|^2 - \|x_{FA_t}^t\|^2, \\
&\le \frac{1}{1 - \delta_{2k}}\|A(x^t - x^\star)\|^2 + \|z_{MD_t}^{t+1}\|^2 - \|x_{FA_t}^t\|^2, &&\text{by RIP} \\
&= \frac{2}{1 - \delta_{2k}}f(x^t) + \|z_{MD_t}^{t+1}\|^2 - \|x_{FA_t}^t\|^2,
\end{aligned}
$$

where the third line follows from the fact that MD_t, FA_t, and CO_t are disjoint sets.

As $f(x^t) > 0$ and $\delta_{2k} < 1 - \frac{1}{2\eta}$, the above inequality implies

$$0 < 2(2\eta - \frac{1}{1 - \delta_{2k}})f(x^t) \le \|z_{MD_t}^{t+1}\|^2 - \|x_{FA_t}^t\|^2.$$

$\square$

Next, we provide a lemma that bounds the function value $f(x^t)$ in terms of missed detections MD_t and also $z_{MD_t}^{t+1}$.

Lemma 16. *Let* $f(x^t) = \frac{1}{2}\|Ax^t - b\|^2$, $b = Ax^*$, $\delta_{2k} < 1 - \frac{1}{2\eta}$ *and* $\eta < 1$. *Then, at each step,*

$$\frac{(1 - \eta)^2}{\eta}\|x_{MD_t}^\star\|^2 \le f(x^t) \le \frac{1}{4\eta(1 - \eta)^2}\|z_{MD_t}^{t+1}\|^2 \tag{9}$$

Proof. Now, using Lemma 2 of [4] with $I = MD_t$, $J = I_t$, $y = A_{MD_t}x_{MD_t}^\star$ we get

$$
\begin{aligned}
f(x^t) &= \tfrac{1}{2}\|Ax^t - b\|^2 \\
&= \tfrac{1}{2}\|A_{I_t}(x^t - x^\star)_{I_t} - A_{MD_t}x_{MD_t}^\star\|^2 \tag{10} \\
&\ge \tfrac{1}{2}\left(1 - \frac{\delta_{2k}}{1 - \delta_k}\right)^2 \|A_{MD_t}x_{MD_t}^\star\|^2 \\
&\ge \tfrac{1}{2}\left(1 - \frac{\delta_{2k}}{1 - \delta_k}\right)^2 (1 - \delta_k)\|x_{MD_t}^\star\|^2 \qquad \text{by RIP} \\
&\ge \tfrac{1}{2}\left(1 - \frac{\delta_{2k}}{1 - \delta_{2k}}\right)^2 (1 - \delta_{2k})\|x_{MD_t}^\star\|^2 \\
&\ge \frac{(1 - 2\delta_{2k})^2}{2(1 - \delta_{2k})}\|x_{MD_t}^\star\|^2 \tag{11}
\end{aligned}
$$

The assumption that $\delta_{2k} < 1 - \frac{1}{2\eta}$ and $\eta < 1$ implies that $\delta_{2k} < 1 - \frac{1}{2\eta} < 1/2$. The function $\alpha \mapsto (1 - 2\alpha)^2/(2(1 - \alpha))$ is decreasing on $[0, 1/2]$ and hence (11) implies

$$f(x^t) \ge \frac{\left(1 - 2(1 - \frac{1}{2\eta})\right)^2}{2(1 - 1 + \frac{1}{2\eta})}\|x_{MD_t}^\star\|^2 = \frac{(1 - \eta)^2}{\eta}\|x_{MD_t}^\star\|^2. \tag{12}$$

Next, using (7) and Cauchy-Schwarz inequality:

$$\|z_{MD_t}^{t+1}\|^2 \ge 4\eta^2 \frac{f(x^t)^2}{\|x_{MD_t}^\star\|^2}. \tag{13}$$

The result now follows using the above equation with (12). $\square$

Lemma 17. *Let* $\delta_{2k} < 1 - \frac{1}{2\eta}$ *and* $1/2 < \eta < 1$. *Then assuming* $f(x^t) > 0$, *at least one new element is found i.e.* $F_t \ne \emptyset$. *Furthermore,* $\|y_{F_t}^{t+1}\|^2 > \frac{l}{k}cf(x^t)$, *where* $c = \min(4\eta(1 - \eta)^2, 2(2\eta - \frac{1}{1 - \delta_{2k}})) > 0$ *is a constant.*

Proof. We consider the following three exhaustive cases:

1. $|F_t| < l$ and $|F_t| < |MD_t|$: Here, we first argue that $F_t \ne \emptyset$. Assuming $\delta_{2k} < 1 - 1/2\eta$, $f(x_t) > 0$ and using Lemma 15,

$$\|z_{MD_t}^{t+1}\| > \|x_{FA_t}^t\|.$$

Also, $|MD_t| = |FA_t|$. Using (3), $z_{FA_t}^{t+1} = x_{FA_t}^t$. Now partial hard-thresholding selects top l elements from z^{t+1}, hence at least one element of $x_{FA_t}^t$ must not have been selected in I_{t+1}(as MD_t should have at least one larger element). Hence F_t and L_t cannot be empty.

11

Let $S \subseteq |MD_t \backslash F_t|$, s.t., $|S| = |F_t| - |MD_t \cap F_t|$. Now,

$$|S \cup (MD_t \cap F_t)| = |F_t|, \quad |(MD_t \backslash F_t) \backslash S| = |MD_t| - |F_t|.$$

Now, as y_{F_t} consists of top F_t elements of $z_{MD_t}^{t+1}$:

$$\|z_{S \cup (MD_t \cap F_t)}^{t+1}\|^2 \leq \|y_{F_t}\|^2. \tag{14}$$

Furthermore, since $|F_t| < l$, hence every element of $z_{MD_t \backslash F_t}^{t+1}$ is smaller in magnitude than every element of $x_{FA_t \backslash L_t}^t$, otherwise that element should have been included in F_t. Furthermore, $|MD_t| - |F_t| = |FA_t| - |L_t| \leq |FA_t \backslash L_t|$. Hence,

$$\|z_{(MD_t \backslash F_t) \backslash S}^{t+1}\|^2 \leq \|x_{FA_t \backslash L_t}^t\|^2 \leq \|x_{FA_t}^t\|^2, \tag{15}$$

Adding (14) and (15), we get:

$$\|z_{MD_t}^{t+1}\|^2 \leq \|y_{F_t}^{t+1}\|^2 + \|x_{FA_t}^t\|^2. \tag{16}$$

Using above equation along with Lemma 15, we get:

$$\|y_{F_t}^{t+1}\|^2 \geq 2 \left(2\eta - \frac{1}{1 - \delta_{2k}} \right) f(x^t). \tag{17}$$

Now, note that if $|F_t| = 0$, then $y_{F_t}^{t+1} = 0$ implying that $f(x^t) = 0$. Hence, at least one new element is added, i.e., $y_{F_t}^{t+1} \neq \emptyset$.

2. $|F_t| = l < |MD_t|$: By definition of $y_{F_t}^{t+1}$:

$$\frac{\|y_{F_t}^{t+1}\|^2}{|F_t|} \geq \frac{\|z_{MD_t}^{t+1}\|^2}{|MD_t|}.$$

Hence, using Lemma 16 and the fact that $|F_t| = l$:

$$\|y_{F_t}^{t+1}\|^2 \geq \frac{l}{|MD_t|} 4\eta(1 - \eta)^2 f(x^t) \geq \frac{l}{k} 4\eta(1 - \eta)^2 f(x^t), \tag{18}$$

as $|MD_t| \leq k$.

3. $|F_t| \geq |MD_t|$: Since, $y_{F_t}^{t+1}$ is the top most elements of z^{t+1}. Hence, assuming $|F_t| \geq |MD_t|$,

$$\|y_{F_t}^{t+1}\|^2 \geq \|z_{MD_t}^{t+1}\|^2.$$

Now, using Lemma 16:

$$\|y_{F_t}^{t+1}\|^2 \geq 4\eta(1 - \eta)^2 f(x^t). \tag{19}$$

We get the lemma by combining bounds for all the three cases, i.e., (17), (18), (19). $\qquad \square$

Now we give a complete proof of Theorem 4.

Proof. We have,

$$f(y^{t+1}) - f(x^t) = (y^{t+1} - x^t)^T A^T A (x^t - x^\star) + 1/2 \|A(y^{t+1} - x^t)\|^2,$$

$$\leq (y^{t+1} - x^t)^T A^T A (x^t - x^\star) + \frac{(1 + \delta_{2l})}{2} (\|y_{F_t}^{t+1}\|^2 + \|x_{L_t}^t\|^2). \tag{20}$$

where the second inequality follows by using the fact that $y_{I_{t+1} \cap I_t}^{t+1} = x_{I_{t+1} \cap I_t}^t$ and using RIP of order $2l$ (since $|\operatorname{supp}(y^{t+1} - x^t)| = |F_t \cup L_t| \leq 2l$).

Since $x_{I_t}^t$ is obtained using least squares,

$$A_{I_t}^T A (x^t - x^\star) = \mathbf{0}.$$

Thus, $A_{L_t}^T A (x^t - x^\star) = \mathbf{0}$, because $L_t \subseteq I_t$. Next, note that

$$y_{F_t}^{t+1} = -\eta A_{F_t}^T A (x^t - x^\star).$$

Hence,

$$f(y^{t+1}) - f(x^t) \le \left(\frac{1 + \delta_{2l}}{2} - \frac{1}{\eta}\right) \|y^{t+1}_{F_t}\|^2 + \frac{1 + \delta_{2l}}{2} \|x^t_{L_t}\|^2. \tag{21}$$

Furthermore, since y^{t+1} is chosen based on the k largest entries in $z^{t+1}_{J_{t+1}}$, we have,

$$\|y^{t+1}_{F_t}\|^2 = \|z^{t+1}_{F_t}\|^2 \ge \|z^{t+1}_{L_t}\|^2 = \|x^t_{L_t}\|^2 .$$

Plugging this into (21), we get:

$$f(y^{t+1}) - f(x^t) \le \left(1 + \delta_{2l} - \frac{1}{\eta}\right) \|y^{t+1}_{F_t}\|^2 .$$

Now, using Lemma 17, $\|y^{t+1}_{F_t}\|^2 \ge \frac{l}{k} c f(x^t)$ and therefore,

$$f(x^{t+1}) - f(x^t) \le f(y^{t+1}) - f(x^t) \le -\alpha \frac{l}{k} f(x^t)$$

where $\alpha = c\left(1 + \delta_{2l} - \frac{1}{\eta}\right) > 0$ since $\eta(1 + \delta_{2l}) < 1$. Hence,

$$f(x^{t+1}) \le (1 - \alpha \frac{l}{k}) f(x^t) \le e^{-\alpha \frac{l}{k}} f(x^t).$$

The above inequality shows that at each iteration OMPR (l) reduces the objective function value by a fixed multiplicative factor. Furthermore, if x^0 is chosen to have entries bounded by 1, then $f(x^0) \le (1 + \delta_{2k})k$. Hence, after $O(\frac{k}{l} \log((1 + \delta_{2k})k/\epsilon))$ iterations, the function value reduces to ϵ, i.e., $f(x^t) \le \epsilon$. $\qquad \square$

B   Proofs related to the LSH Section

Lemma 18. *Let $\|x\| = 1$ for all points x in our database. Let x^* be the nearest neighbor to r in L_2 distance metric, and let $r^T x^* \ge c > 0$. Then, a $(1 + \alpha\epsilon)$-nearest neighbor to r is also a $(1 - \epsilon)$-similar neighbor to r, where $\alpha \le \frac{2c}{1 + r^T r - 2c}$.*

Proof. Let x' be a $(1 + \alpha\epsilon)$-nearest neighbor to r, then:

$$\|x' - r\|^2 \le (1 + \alpha\epsilon)\|x^* - r\|^2.$$

Using $\|x'\| = \|x^*\| = 1$ and simplifying, we get:

$$r^T x' \ge (1 - \epsilon)r^T x^* + (\alpha + 1)\epsilon r^T x^* - \frac{\alpha\epsilon}{2}(1 + r^T r),$$

$$\ge (1 - \epsilon)r^T x^* + ((\alpha + 1)c - \frac{\alpha}{2}(1 + r^T r))\epsilon.$$

Hence, x' is a $(1 - \epsilon)$-approximate similar neighbor to r if:

$$(\alpha + 1)c \ge \frac{\alpha}{2}(1 + r^T r).$$

The result follows after simplification. $\qquad \square$

We now provide a proof of Theorem 7.

Proof. Let us first consider a single step of OMPR . Now, similar to Lemma 15, we can show that if $\delta_{2k} < 1/4 - \gamma$ and $\eta = 1 - \gamma, \gamma > 0$, then $\|z^{t+1}_{MD_t}\|^2 > \frac{3}{2}\|x^t_{FA_t}\|^2$. Setting $\epsilon = 1 - \sqrt{\frac{2}{3}}$, implies that $(1 - \epsilon) \max |z^{t+1}_{MD_t}| \ge \min |x^t_{FA_t}|$, i.e., a $(1 - \epsilon)$-similar neighbor to $\max |z^{t+1}_{MD_t}|$ will still lead to a constant decrease in the objective function.

So, the goal is to ensure that with probability $1 - \delta$, $\delta > 0$, for all the $O(k)$ iterations, our LSH method returns at least a $(1 - \epsilon)$-similar neighbor to $\max |z_{MD_t}^{t+1}|$ where $\epsilon = 1 - \sqrt{\frac{2}{3}}$. To this end, we need to ensure that at each step t, LSH finds at least a $(1 - \epsilon)$-similar neighbor to $\max |z_{MD_t}^{t+1}|$ with probability at least $1 - \delta/k$. Using Lemma 18, we need to find a $(1 + \alpha\epsilon)$-nearest neighbor to $\max |z_{MD_t}^{t+1}|$, where

$$\alpha \le \frac{2c}{1 + \boldsymbol{r}^T \boldsymbol{r} - 2c},$$

and $\boldsymbol{r}^T x^* \ge c$. Using Lemma 17, $\alpha = O(1/k)$. Hence the result now follows using Theorem 6 (main text).   □

C   Extension to Noisy Case

In this section, we consider the noisy case in which our objective function is $f(x) = \frac{1}{2}\|Ax - b\|^2$, where $b = Ax^* + e$ and $e \in \mathbb{R}^m$ is the "noise" vector.

Let I_t denote the support set of x^t and $I^\star$ be the support set of $x^\star$. Define the sets

$$
\begin{aligned}
FA_t &= I_t \backslash I^\star & \text{(false alarms)} \\
MD_t &= I^\star \backslash I_t & \text{(missed detections)} \\
CO_t &= I_t \cap I^\star & \text{(correct detections)} .
\end{aligned}
$$

Lemma 19. *Let* $f(x^t) \ge \frac{C}{2}\|e\|^2$ *and* $\delta_{2k} < 1 - \frac{1}{2D\eta}$, *where* $D = \frac{C - \sqrt{C}}{(\sqrt{C}+1)^2}$. *Then,*

$$\|z_{MD_t}^{t+1}\|^2 - \|x_{FA_t}^t\|^2 \ge cf(x^t),$$

where $c = 2\frac{(\sqrt{C}+1)^2}{C}(2\eta D - \frac{1}{1-\delta_{2k}}) > 0$.

Proof. Since $x_{I_t}^t$ is the solution to the least squares problem $\min_x \|A_{I_t} x - b\|^2$,

$$A_{I_t}^T (A_{I_t} x_{I_t}^t - b) = 0. \tag{22}$$

Now, note that

$$
\begin{aligned}
f(x^t) &= \frac{1}{2}\|A_{I_t} x_{I_t}^t - b\|^2, \\
&= \frac{1}{2}((x_{I_t}^t)^T A_{I_t}^T (A_{I_t} x_{I_t}^t - b) - b^T (A_{I_t} x_{I_t}^t - b)), \\
&= -\frac{1}{2} b^T (A_{I_t} x_{I_t}^t - b), \\
&= -\frac{1}{2}(x_{MD_t}^*)^T A_{MD_t}^T (A_{I_t} x_{I_t}^t - b) - \frac{1}{2} e^T (A_{I_t} x_{I_t}^t - b), \\
&= \frac{1}{2\eta}(x_{MD_t}^\star)^T z_{MD_t}^{t+1} - \frac{1}{2} e^T (A_{I_t} x_{I_t}^t - b), \tag{23}
\end{aligned}
$$

where the third equality follows from (22).

Now,

$$
\begin{aligned}
\|x_{MD_t}^\star - z_{MD_t}^{t+1}\|^2 &= \|x_{MD_t}^\star\|^2 + \|z_{MD_t}^{t+1}\|^2 - 2(x_{MD_t}^\star)^T z_{MD_t}^{t+1} \\
&= \|x_{MD_t}^\star\|^2 + \|z_{MD_t}^{t+1}\|^2 - 4\eta(f(x^t) + \frac{1}{2} e^T (A_{I_t} x_{I_t}^t - b)) \tag{24}
\end{aligned}
$$

14

So,

$$0 \leq \|x^\star_{MD_t}\|^2 + \|x^t_{FA_t}\|^2 - \|x^t_{FA_t}\|^2 + \|z^{t+1}_{MD_t}\|^2 - 4\eta(f(x^t) + \frac{1}{2}e^T(A_{I_t}x^t_{I_t} - b)),$$

$$\leq \|x^\star_{MD_t}\|^2 + \|x^t_{FA_t}\|^2 + \|x^t_{CO_t} - x^\star_{CO_t}\|^2 - \|x^t_{FA_t}\|^2 + \|z^{t+1}_{MD_t}\|^2 - 4\eta(f(x^t) + \frac{1}{2}e^T(A_{I_t}x^t_{I_t} - b)),$$

$$\leq \|x^t - x^*\|^2 - \|x^t_{FA_t}\|^2 + \|z^{t+1}_{MD_t}\|^2 - 4\eta(f(x^t) + \frac{1}{2}e^T(A_{I_t}x^t_{I_t} - b)),$$

$$\leq \frac{1}{1-\delta_{2k}}\|A(x^t - x^*)\|^2 - \|x^t_{FA_t}\|^2 + \|z^{t+1}_{MD_t}\|^2 - 4\eta(f(x^t) + \frac{1}{2}e^T(A_{I_t}x^t_{I_t} - b)),$$

$$= \frac{1}{1-\delta_{2k}}\|A(x^t - x^*)\|^2 - \|x^t_{FA_t}\|^2 + \|z^{t+1}_{MD_t}\|^2 - 4\eta(1 - \frac{1}{\sqrt{C}})f(x^t).$$

Now, by assumption: $f(x^t) \geq \frac{C}{2}\|e\|^2$. Hence,

$$\|A(x^t - x^*)\| \leq \|A(x^t - x^*) - e\| + \|e\|,$$

$$\|A(x^t - x^*)\|^2 \leq 2(1 + \frac{1}{\sqrt{C}})^2 f(x^t).$$

Hence,

$$2\left(2\eta(1 - \frac{1}{\sqrt{C}}) - \frac{1}{1-\delta_{2k}}(1 + \frac{1}{\sqrt{C}})^2\right) f(x^t) \leq \|z^{t+1}_{MD_t}\|^2 - \|x^t_{FA_t}\|^2$$

Now, by assumption $\delta_{2k} < 1 - \frac{1}{2D\eta}$, where $D = \frac{(\sqrt{C}+1)^2}{C - \sqrt{C}}$. Hence, $c = 2\frac{(\sqrt{C}+1)^2}{C}(2\eta D - \frac{1}{1-\delta_{2k}}) > 0$. $\qquad\square$

Next, we provide a lemma that bounds the function value $f(x^t)$ in terms of missed detection MD_t and also $z^{t+1}_{MD_t}$.

Lemma 20. *Let* $f(x^t) = \frac{1}{2}\|Ax^t - b\|^2 \geq \frac{C}{2}\|e\|^2$, $b = Ax^* + e$, $\delta_{2k} < 1 - \frac{1}{2D\eta}$ *and* $D = \frac{C - \sqrt{C}}{(\sqrt{C}+1)^2}$. *Then, at each step,*

$$\frac{(1-\eta)^2 C}{\eta(\sqrt{C}+1)^2}\|x^\star_{MD_t}\|^2 \leq f(x^t) \leq \frac{1}{4\eta(1-\eta)^2}\frac{(\sqrt{C}+1)^2}{(\sqrt{C}-1)^2}\|z^{t+1}_{MD_t}\|^2 \qquad (25)$$

Proof. First we lower bound $f(x^t)$:

$$\sqrt{f(x^t)} = \frac{1}{\sqrt{2}}\|Ax^t - Ax^* - e\|,$$

$$\geq \frac{1}{\sqrt{2}}\left(\|Ax^t - Ax^*\| - \|e\|\right),$$

$$\geq \frac{1}{\sqrt{2}}\left(\min_{x\,:\,x_{\bar{I}_t}=0}\|Ax - Ax^*\| - \|e\|\right),$$

$$\geq \frac{1}{\sqrt{2}}\left(\frac{(1-2\delta_{2k})}{\sqrt{(1-\delta_{2k})}}\|x^\star_{MD_t}\| - \|e\|\right),$$

where last equality follows from Lemma 16. Using the above inequality with $f(x^t) \geq \frac{C}{2}\|e\|^2$, we get:

$$f(x^t) \geq \frac{(1-2\delta_{2k})^2 C}{2(1-\delta_{2k})(\sqrt{C}+1)^2}\|x^\star_{MD_t}\|^2. \qquad (26)$$

The assumption that $\delta_{2k} < 1 - \frac{1}{2D\eta}$ and $D\eta < 1$ implies that $\delta_{2k} < 1 - \frac{1}{2D\eta} < 1/2$. The function $\alpha \mapsto (1 - 2\alpha)^2/(2(1-\alpha))$ is decreasing on $[0, 1/2]$ and hence the above equation implies

$$f(x^t) \geq \frac{(1-D\eta)^2}{D\eta}\|x^\star_{MD_t}\|^2. \qquad (27)$$

15

Now, we upper bound $f(x^t)$. Using definition of $f(x^t)$:

$$\frac{1}{2\eta}(x^\star_{MD_t})^T z^{t+1}_{MD_t} = f(x^t) + \frac{1}{2}e^T(A_{I_t} x^t_{I_t} - b).$$

Now, using Cauchy-Schwarz and $f(x^t) \geq \frac{C}{2}\|e\|^2$,

$$\left|e^T(A_{I_t} x^t_{I_t} - b)\right| \leq \|e\|\|A_{I_t} x^t_{I_t} - b\| \leq \frac{2}{\sqrt{C}}f(x^t).$$

Hence,

$$\frac{1}{2\eta}\|x^\star_{MD_t}\|\|z^{t+1}_{MD_t}\| \geq \frac{1}{2\eta}(x^\star_{MD_t})^T z^{t+1}_{MD_t} \geq (1 - \frac{1}{\sqrt{C}})f(x^t).$$

That is,

$$\|z^{t+1}_{MD_t}\|^2 \geq 4\eta^2\left(1 - \frac{1}{\sqrt{C}}\right)^2 \frac{f(x^t)^2}{\|x^\star_{MD_t}\|^2} \geq 4\eta(1 - D\eta)^2 \frac{(\sqrt{C}-1)^2}{CD}f(x^t), \qquad (28)$$

where the second inequality follows from (27). $\qquad\square$

Next, we present the following lemma that shows "enough" progress at each step:

Lemma 21. *Let $f(x^t) \geq \frac{C}{2}\|e\|^2$, $\eta < 1$ and $\delta_{2k} < 1 - \frac{1}{2D\eta}$, where $D = 1 - \frac{1}{\sqrt{C}-1}$. Then at least one new element is found i.e. $F_t \neq \emptyset$. Furthermore, $\|y^{t+1}_{F_t}\| > \frac{l}{k}\alpha f(x^t)$, where $\alpha = \min(4\eta(1 - D\eta)^2 \frac{(\sqrt{C}-1)^2}{CD}, 2\frac{(\sqrt{C}+1)^2}{C}(2\eta D - \frac{1}{1-\delta_{2k}})) > 0$ is a constant.*

Proof. As for the exact case, we analyse the following three exhaustive cases:

1. $|F_t| < l$ and $|F_t| < |MD_t|$: Here we use the exactly similar argument as the exact case to obtain the following inequality (see (16)):
$$\|z^{t+1}_{MD_t}\|^2 \leq \|y^{t+1}_{F_t}\|^2 + \|x^t_{FA_t}\|^2. \qquad (29)$$
Using Lemma 19, we get:
$$\|y^{t+1}_{F_t}\|^2 \geq cf(x^t), \qquad (30)$$
where c is as defined in Lemma 19. Now, note that if $|F_t| = 0$, then $y^{t+1}_{F_t} = 0$ implying that $f(x^t) = 0$. Hence, at least one new element is added, i.e., $y^{t+1}_{F_t} \neq \emptyset$.

2. $|F_t| = l < |MD_t|$: By definition of $y^{t+1}_{F_t}$:
$$\frac{\|y^{t+1}_{F_t}\|^2}{|F_t|} \geq \frac{\|z^{t+1}_{MD_t}\|^2}{|MD_t|}.$$
Hence, using Lemma 20 and the fact that $|F_t| = l$:
$$\|y^{t+1}_{F_t}\|^2 \geq \frac{l}{|MD_t|}4\eta(1 - D\eta)^2\frac{(\sqrt{C}-1)^2}{CD}f(x^t) \geq \frac{l}{k}4\eta(1 - D\eta)^2\frac{(\sqrt{C}-1)^2}{CD}f(x^t), \qquad (31)$$
as $|MD_t| \leq k$.

3. $|F_t| \geq |MD_t|$: Since, $y^{t+1}_{F_t}$ is the top most elements of z^{t+1}. Hence, assuming $|F_t| \geq |MD_t|$,
$$\|y^{t+1}_{F_t}\|^2 \geq \|z^{t+1}_{MD_t}\|^2.$$
Now, using Lemma 20:
$$\|y^{t+1}_{F_t}\|^2 \geq 4\eta(1 - D\eta)^2\frac{(\sqrt{C}-1)^2}{CD}f(x^t). \qquad (32)$$

We get the lemma by combining bounds for all the three cases, i.e., (30), (31), (32). $\qquad\square$

Now, we provide a proof of Theorem 2.

Proof. We have,

$$f(y^{t+1}) - f(x^t) = (y^{t+1} - x^t)^T A^T (Ax^t - b) + 1/2\|A(y^{t+1} - x^t)\|^2,$$

$$\leq (y^{t+1} - x^t)^T A^T (Ax^t - b) + \frac{(1 + \delta_{2l})}{2} (\|y_{F_t}^{t+1}\|^2 + \|x_{L_t}^t\|^2). \tag{33}$$

where the second inequality follows by using the fact that $y_{I_{t+1} \cap I_t}^{t+1} = x_{I_{t+1} \cap I_t}^t$ and using RIP of order $2l$ (since $|\operatorname{supp}(y^{t+1} - x^t)| = |F_t \cup L_t| \leq 2l$).

Since $x_{I_t}^t$ is obtained using least squares,

$$A_{I_t}^T (Ax^t - b) = \mathbf{0}.$$

That is, $A_{L_t}^T (Ax^t - b) = \mathbf{0}$, because $L_t \subseteq I_t$. Next, note that

$$y_{F_t}^{t+1} = -\eta A_{F_t}^T (Ax^t - b).$$

Hence,

$$f(y^{t+1}) - f(x^t) \leq \left(\frac{1 + \delta_{2l}}{2} - \frac{1}{\eta} \right) \|y_{F_t}^{t+1}\|^2 + \frac{1 + \delta_{2l}}{2} \|x_{L_t}^t\|^2. \tag{34}$$

Furthermore, since y^{t+1} is chosen based on largest entries in $z_{J_{t+1}}^{t+1}$, we have,

$$\|y_{F_t}^{t+1}\|^2 = \|z_{F_t}^{t+1}\|^2 \geq \|z_{L_t}^{t+1}\|^2 = \|x_{L_t}^t\|^2 .$$

Plugging this into (34), we get:

$$f(y^{t+1}) - f(x^t) \leq \left(1 + \delta_{2l} - \frac{1}{\eta} \right) \|y_{F_t}^{t+1}\|^2 .$$

Now, using Lemma 21, $\|y_{F_t}^{t+1}\|^2 \geq \alpha f(x^t) > 0$ and therefore,

$$f(x^{t+1}) - f(x^t) \leq f(y^{t+1}) - f(x^t)$$

$$\leq -c' \frac{l}{k} f(x^t) ,$$

where $c' = \frac{1 - \eta(1 + \delta_{2l})}{\eta(1 + \delta_{2l})} \alpha > 0$ since $\eta(1 + \delta_{2l}) < 1$. The above inequality shows that at each iteration OMPR (l) reduces the objective function value by a fixed multiplicative factor. Furthermore, if x^0 is chosen to have entries bounded by 1, then $f(x^0) \leq O((1 + \delta_{2k})k + \|e\|^2)$. Hence, after $O(\frac{k}{l} \log((k + \|e\|^2)/\epsilon))$ iterations, the function value reduces to $C\|e\|^2/2 + \epsilon$. □

D   Analysis of 2-stage Algorithms

In this section, we consider the family of two-stage hard thresholding algorithms (see Algorithm 3) introduced by [17].

We now provide a simple analysis for the general two-stage hard thresholding algorithms. We first present a few technical lemmas that we will need for our proof.

Lemma 22. *Let $b = Ax^\star$, where $I^* = \operatorname{supp}(x^\star)$. Also, let $x = \operatorname{argmin}_{\operatorname{supp}(x)=I} \|Ax - b\|^2$. Then,*

$$\sqrt{\|(x - x^\star)_{I \cap I^*}\|^2 + \|x_{I \setminus I^*}\|^2} = \|(x - x^\star)_I\| \leq \frac{\delta_{|I \cup I^*|}}{\sqrt{1 - \delta_{|I \cup I^*|}^2}} \|x_{I^* \setminus I}^\star\|$$

Algorithm 3 Two-stage(l)

1: **Input:** matrix A, vector b, sparsity level k
2: Initialize x^1
3: **for** $t = 1$ **to** T **do**
4:    $\text{top}_{t+1} \leftarrow$ indices of top l elements of $|A^T(Ax^t - b)|$
5:    $J_{t+1} \leftarrow I_t \cup \text{top}_{t+1}$
6:    $z_{J_{t+1}}^{t+1} \leftarrow A_{J_{t+1}} \backslash b$, $z_{\bar{J}_{t+1}}^{t+1} \leftarrow \mathbf{0}$
7:    $y^{t+1} \leftarrow H_k\left(z^{t+1}\right)$
8:    $I_{t+1} \leftarrow \text{supp}(y^{t+1})$
9:    $x_{I_{t+1}}^{t+1} \leftarrow A_{I_{t+1}} \backslash b$, $x_{\bar{I}_{t+1}}^{t+1} \leftarrow \mathbf{0}$
10: **end for**

Proof. A similar inequality appears in [10] and we rewrite the proof here. Since x_I is the solution to $\min_u \|A_I u - b\|^2$,
$$A_I^T(A_I x_I - b) = 0. \tag{35}$$
In the exact case, $b = Ax^*$. Hence,
$$\|(x - x^\star)_I\|^2 = [(x - x^\star)_I \ \ 0]^T \begin{bmatrix} (x - x^\star)_I \\ -x_{I^* \backslash I}^\star \end{bmatrix} \tag{36}$$
Now, using (35):
$$0 = [(x - x^\star)_I \ \ 0]^T A_G^T A_G \begin{bmatrix} (x - x^\star)_I \\ -x_{I^* \backslash I}^\star \end{bmatrix}, \tag{37}$$
where $G = [I \ \ I^* \backslash I]$. Subtracting (37) from (36) we get,
$$\|(x - x^\star)_I\|^2 = [(x - x^\star)_I \ \ 0]^T (I - A_G^T A_G) \begin{bmatrix} (x - x^\star)_I \\ -x_{I^* \backslash I}^\star \end{bmatrix},$$
$$\leq \delta_{2k} \|(x - x^\star)_I\| \sqrt{\|(x - x^\star)_I\|^2 + \|x_{I^* \backslash I}^\star\|^2}, \tag{38}$$
where the second inequality follows using Lemma 13. Lemma follows by just rearranging terms now. $\qquad\square$

We now present our main theroem and its proof for two-stage thresholding algorithms.
Theorem 23. *Suppose the vector $x^\star \in \mathbb{R}^n$ is k-sparse and binary. Then Two-stage(l) recovers $x^\star$ from measurements $b = Ax^\star$ in $O(k)$ iterations provided:*
$$\delta_{2k+l} \leq .35$$

Proof. As z^{t+1} is the least squares solution over support set J_{t+1}, hence:
$$f(z^{t+1}) - f(x^t) \leq f(s^{t+1}) - f(x^t), \tag{39}$$
where $s_{J_{t+1}}^{t+1} = (x^t - \eta A^T(Ax^t - b))_{J_{t+1}}$, $\eta = \frac{1}{1+\delta_l}$ and $s_{\bar{J}_{t+1}}^{t+1} = 0$.

Now,
$$f(s^{t+1}) - f(x^t) = (s^{t+1} - x^t)^T A^T(Ax^t - b) + \frac{1}{2}\|As^{t+1} - Ax^t\|^2. \tag{40}$$
Now, as x^t is the least squares solution over I_t. Hence, $A_{I_t}^T(Ax^t - b) = 0$. Hence,
$$(s^{t+1} - x^t)_{I_t} = 0, \quad (s^{t+1} - x^t)_{\text{top}_{t+1}} = -\eta A_{\text{top}_{t+1}}^T(Ax^t - b), \quad (s^{t+1} - x^t)_{\bar{J}_{t+1}} = 0. \tag{41}$$
Using (40) and (41):
$$f(s^{t+1}) - f(x^t) = -\eta\|A_{\text{top}_{t+1}}^T(Ax^t - b)\|^2 + \frac{\eta^2}{2}\|A_{\text{top}_{t+1}}A_{\text{top}_{t+1}}^T(Ax^t - b)\|^2,$$
$$\leq -\eta\|A_{\text{top}_{t+1}}^T(Ax^t - b)\|^2 + \frac{\eta^2(1 + \delta_l)}{2}\|A_{\text{top}_{t+1}}^T(Ax^t - b)\|^2,$$
$$= -\frac{\eta}{2}\|A_{\text{top}_{t+1}}^T(Ax^t - b)\|^2. \tag{42}$$

Now, let MD_t be the set of missed detections, i.e., $MD_t = I^* \backslash I^t$. Then, by definition of top_{t+1}:

$$\|A_{\text{top}_{t+1}}^T (Ax^t - b)\|^2 \geq \min\left(1, \frac{l}{|MD_t|}\right) \|A_{MD_t}^T (Ax^t - b)\|^2. \tag{43}$$

Furthermore,

$$\begin{aligned}
\|A_{MD_t}^T (Ax^t - b)\| &= \|A_{MD_t}^T A_{MD_t} x_{MD_t}^* - A_{MD_t}^T A_{I_t} (x^t - x^*)_{I_t}\|, \\
&\geq \|A_{MD_t}^T A_{MD_t} x_{MD_t}^*\| - \|A_{MD_t}^T A_{I_t} (x^t - x^*)_{I_t}\|, \\
&\geq (1 - \delta_k)\|x_{MD_t}^*\| - \frac{\delta_{2k}^2}{\sqrt{1 - \delta_{2k}^2}} \|x_{MD_t}^*\|,
\end{aligned} \tag{44}$$

where last inequality follows using Lemma 14 and Lemma 22.

Hence, using (42), (43), and (44):

$$f(z^{t+1}) - f(x^t) \leq f(s^{t+1}) - f(x^t) \leq -\frac{1}{2(1 + \delta_l)} \min\left(1, \frac{l}{|MD_t|}\right) \left(1 - \delta_k - \frac{\delta_{2k}^2}{\sqrt{1 - \delta_{2k}^2}}\right)^2 \|x_{MD_t}^*\|^2. \tag{45}$$

Next, we upper bound increase in the objective function by removing l elements from z^{t+1}.

$$\begin{aligned}
f(y^{t+1}) - f(z^{t+1}) &= (y^{t+1} - z^{t+1})^T A^T (Az^{t+1} - b) + \frac{1}{2}\|Ay^{t+1} - Az^{t+1}\|^2, \\
&= \frac{1}{2}\|Ay^{t+1} - Az^{t+1}\|^2, \\
&\leq \frac{1 + \delta_l}{2}\|z_{J_{t+1} \backslash I_{t+1}}^{t+1}\|^2,
\end{aligned} \tag{46}$$

where the second equation follows as z^{t+1} is a least squares solution, and both y^{t+1}, z^{t+1}'s support is a subset of J_{t+1}. The third equation follows from RIP and the fact that $z_{I_{t+1}}^{t+1} = y_{I_{t+1}}^{t+1}$.

Now, using Lemma 22:

$$\|z_{J_{t+1} \backslash I^*}^{t+1}\|^2 \leq \frac{\delta_{2k+l}^2}{1 - \delta_{2k+l}^2} \|x_{I^* \backslash J_{t+1}}^*\|^2. \tag{47}$$

Furthermore, $|J_{t+1} \backslash I_{t+1}| = l \leq |J_{t+1} \backslash I^*|$. Hence, by definition of I_{t+1},

$$\|z_{J_{t+1} \backslash I_{t+1}}^{t+1}\|^2 \leq \frac{l}{|J_{t+1} \backslash I^*|} \|z_{J_{t+1} \backslash I^*}^{t+1}\|^2.$$

Using above equation and (47), we get:

$$\|z_{J_{t+1} \backslash I_{t+1}}^{t+1}\|^2 \leq \frac{l}{|J_{t+1} \backslash I^*|} \frac{\delta_{2k+l}^2}{1 - \delta_{2k+l}^2} \|x_{I^* \backslash J_{t+1}}^*\|^2, \tag{48}$$

Also, $|J_{t+1} \backslash I^*| = l + |I^* \backslash J_{t+1}| \leq l + |MD_t|$. Using (46), (48), and the fact that $f(x^{t+1}) \leq f(y^{t+1})$ and each $x_{I^*}^* = 1$:

$$f(x^{t+1}) - f(z^{t+1}) \leq \frac{l}{l + |I^* \backslash J_{t+1}|} \frac{1 + \delta_l}{2} \frac{\delta_{2k+l}^2}{1 - \delta_{2k+l}^2} |I^* \backslash J_{t+1}|. \tag{49}$$

Adding (45) and (49), we get:

$$f(x^{t+1}) - f(x^t) \leq -\frac{1}{2(1 + \delta_l)} \left(\min\left(|MD_t|, l\right) \left(1 - \delta_k - \frac{\delta_{2k}^2}{\sqrt{1 - \delta_{2k}^2}}\right)^2 - \frac{l \cdot |I^* \backslash J_{t+1}|}{l + |I^* \backslash J_{t+1}|} \frac{(1 + \delta_l)^2 \delta_{2k+l}^2}{1 - \delta_{2k+l}^2}\right).$$

$$\tag{50}$$

19

Now, $\frac{l\cdot|I^*\setminus J_{t+1}|}{l+|I^*\setminus J_{t+1}|} \leq \min(l, |I^*\setminus J_{t+1}|) \leq \min(l, |MD_t|)$.

Hence,

$$f(x^{t+1}) - f(x^t) \leq -\frac{\min(l, |MD_t|)}{2(1+\delta_l)}\left(\left(1 - \delta_k - \frac{\delta_{2k}^2}{\sqrt{1-\delta_{2k}^2}}\right)^2 - \frac{(1+\delta_l)^2\delta_{2k+l}^2}{1-\delta_{2k+l}^2}\right). \tag{51}$$

Now consider:

$$\left(\left(1 - \delta_k - \frac{\delta_{2k}^2}{\sqrt{1-\delta_{2k}^2}}\right)^2 - \frac{(1+\delta_l)^2\delta_{2k+l}^2}{1-\delta_{2k+l}^2}\right) \geq \frac{1}{1-\delta_{2k+l}^2}\left(((1-\delta_{2k+l})\sqrt{1-\delta_{2k+l}^2} - \delta_{2k+l}^2)^2 - (1+\delta_{2k+l})^2\delta_{2k+l}^2\right),$$

$$> 0.01, \tag{52}$$

where the second inequality follows by substituting $\delta_{2k+1} \leq .35$.

Hence, using (51) and (52), we have:

$$f(x^{t+1}) \leq f(x^t) - \min(l, |MD_t|) \cdot 0.0001. \tag{53}$$

The above equation guarantees convergence to the optima in at least $O(k)$ steps although faster convergence can be shown for larger k. □

Corollary 24. *Cosamp converges to the optima provided*

$$\delta_{4k} \leq 0.35.$$

Corollary 25. *Subspace-Pursuit converges to the optima provided*

$$\delta_{3k} \leq 0.35.$$

Note that CoSamp's analysis given by [19] requires $\delta_{4k} \leq 0.1$ while Subspace pursuit's analysis given by [4] requires $\delta_{3k} \leq 0.205$. Note that our generic analysis provides significantly better guarantees for both the methods.