A. Proofs of UFF section

In this section we will give proofs for theorems in the UFF section and introduce a robust version of UFF when the measurements have adversarial noise.

Proof of Theorem 2. Consider $k+1$ distinct elements of $\mathcal{F}$: $B_0, B_1, \cdots, B_k$. Let us define the bad event E as

$$E = 1_{\left\{B_0 \subseteq \bigcup_{i=1}^{k} B_i\right\}}.$$

The cardinality of $\bigcup_{i=1}^{k} B_i$ is at most kd. Since the elements of B_0 are chosen independently and uniformly at random from $[m]$, we have:

$$\mathbb{P}[E] \le \frac{\binom{kd}{d}}{\binom{m}{d}} \le \left(\frac{kd}{m-d+1}\right)^d.$$

The total number of choices for the sets $B_0, B_1, \cdots, B_k$ is $n\binom{n-1}{k}$. Using union bound, the probability that Algorithm 1 does not return a (d,k)-UFF is

$$\begin{aligned}\mathbb{P}[\mathrm{Err}] &\le n\binom{n-1}{k}\left(\frac{kd}{m-d+1}\right)^d \\ &\le n\left(\frac{e(n-1)}{k}\right)^k\left(\frac{kd}{m-d+1}\right)^d \\ &\le n^{2k}e^k\left(\frac{1}{9}\right)^{k\log\left(\frac{3n}{\delta}\right)} \\ &= n^{2k}e^k n^{-k\log 9}3^{-k\log 9}\delta^{k\log 9} < \delta.\end{aligned}$$

This finishes the proof. □

Proof of Theorem 3. Let $S^* = \mathrm{supp}(\boldsymbol{x}^*)$. We know that $|S^*| \le k$. Wlog, assume that non-zero elements are in the first k dimensions of $\boldsymbol{x}^*$, i.e., $S^* = \{1, 2, \cdots, |S^*|\}$.

Proof of $\widehat{S} \subseteq S^*$: Consider any $\ell \notin S^*$. As, A is constructed from $\mathcal{F}$ which is (d,k)-UFF (see Definition 1):

$$B_l \nsubseteq B_1 \cup B_2 \cup \cdots \cup B_{|S^*|}.$$

Therefore, $\exists 1 \le i' \le m$ s.t. $i' \in B_\ell$ and $i' \notin B_j \ \forall j \in S^*$. Furthermore, $b_{i'} = \mathbb{1}_{\{\sum_{j:i' \in B_j} x_j^* > 0\}} = 0$. Therefore, $\min_{i \in B_\ell} b_i = 0$, i.e., $\ell \notin \widehat{S}$ (see Step 4 of Algorithm 2), and it follows that $\widehat{S} \subseteq S^*$.

Proof of $S^* \subseteq \widehat{S}$: Now consider any $\ell \in S^*$. For all $i \in B_\ell$, we have: $\quad b_i = \mathbb{1}_{\{\sum_{j:i \in B_j} x_j^* > 0\}} \ge \mathbb{1}_{\{\boldsymbol{x}_\ell^* > 0\}} = 1$. Therefore, $\min_{i \in B_\ell} b_i > 0$, and by Step 5 of Algorithm 2: $\ell \in \widehat{S}$. Hence, $S^* \subseteq \widehat{S}$. □

In the presence of arbitrary adversarial noise, the measurements no longer satisfy (3) but are given by

$$\boldsymbol{b} = \mathrm{Sign}\left(A\boldsymbol{x}^* + \boldsymbol{\eta}\right) \tag{5}$$

where $\boldsymbol{\eta} \in \mathbb{R}^m$ is a sparse vector of outliers and $\|\boldsymbol{\eta}\|_0$ is the number of adversarial errors. In the case of adversarial errors, we use a (d,k,ϵ)-UFF to construct the measurement matrix as in (2) and the following algorithm to reconstruct $\boldsymbol{x}^*$.

Algorithm 8 Support recovery algorithm when A is constructed from a (d,k,ϵ)-UFF

---

input A : measurement matrix, ϵ : robustness parameter, $\boldsymbol{b}$: measurement vector ($\boldsymbol{b} = \mathrm{sign}(A\boldsymbol{x}^* + \boldsymbol{\eta})$)

1: $\widehat{S} \leftarrow \emptyset$

2: **for** $j = 1, \cdots, n$ **do**

3:   **if** $|\mathrm{supp}(\boldsymbol{b}) \cap B_j| > \frac{|B_j|}{2}$ **then**

4:     $\widehat{S} \leftarrow \widehat{S} \cup \{j\}$

5:   **end if**

6: **end for**

output $\widehat{S}$

Theorem 8 shows that Algorithm 8 recovers $\boldsymbol{x}^*$ even in the presence of at most $\left(\frac{1}{2} - \epsilon\right) d$ adversarial errors.

Theorem 8. *Suppose $\boldsymbol{x}^* \in \mathbb{R}^n_{\ge 0}$ is a vector of non-negative elements s.t. $\|\boldsymbol{x}^*\|_0 \le k$, A is a sensing matrix constructed according to (2) and the measurements are according to (5). Suppose further that the underlying UFF is a (d,k,ϵ)-UFF and there are up to $\left(\frac{1}{2} - \epsilon\right) d$ adversarial errors in the measurement (i.e., $\|\boldsymbol{\eta}\|_0 \le \left(\frac{1}{2} - \epsilon\right) d$ where $\boldsymbol{\eta}$ is as in (5)). Then, the set $\widehat{S}$ returned by Algorithm 8 satisfies: $\widehat{S} = \mathrm{supp}(\boldsymbol{x}^*)$.*

Proof. The proof of this theorem is along lines of the proof for Theorem 3. Let $S^* = \mathrm{supp}(\boldsymbol{x}^*)$. We know that $|S^*| \le k$. Wlog, assume that non-zero elements are in the first k dimensions of $\boldsymbol{x}^*$, i.e., $S^* = \{1, 2, \cdots, |S^*|\}$.

We show $\widehat{S} = S^*$, by first proving $\widehat{S} \subseteq S^*$ and then $S^* \subseteq \widehat{S}$.

Proof of $\widehat{S} \subseteq S^*$: Consider any $\ell \notin S^*$. Since A is constructed from $\mathcal{F}$ which is (d,k,ϵ)-UFF (see Definition 3):

$$\left|B_l \cap \left(B_1 \cup B_2 \cup \cdots \cup B_{|S^*|}\right)\right| < \epsilon\,|B_l| = \epsilon d.$$

Since there are at most $\left(\frac{1}{2} - \epsilon\right) d$ adversarial errors, we have

$$\begin{aligned}|\mathrm{supp}(\boldsymbol{b}) \cap B_\ell| &< \epsilon d + \left(\frac{1}{2} - \epsilon\right) d \\ &= \frac{d}{2} = \frac{|B_\ell|}{2}\end{aligned}$$

So, by Step 4 of Algorithm 8, we have $\ell \notin \widehat{S}$. Hence, $\widehat{S} \subseteq S^*$.

Proof of $S^* \subseteq \widehat{S}$: Now consider any $\ell \in S^*$. For every $i \in B_\ell \setminus \text{supp}(\boldsymbol{\eta})$, we have:

$$b_i = \mathbb{1}_{\{\sum_{j: i \in B_j} x_j^* > 0\}} \geq \mathbb{1}_{\{\boldsymbol{x}_\ell^* > 0\}} = 1.$$

So, $|\text{supp}(\boldsymbol{b}) \cap B_\ell| > (1-\epsilon)\, d - \left(\frac{1}{2} - \epsilon\right) d = \frac{d}{2} = \frac{|B_\ell|}{2}$ and by Step 5 of Algorithm 8: $\ell \in \widehat{S}$. Hence, $S^* \subseteq \widehat{S}$. □

B. Proofs of Expanders section

In this section we will prove Theorem 5 for which we need the following lemma.

Lemma 2. *With the sensing matrix A constructed as in section 3.2.2 and $\boldsymbol{b} = sign(A\boldsymbol{x}^*)$ where $\boldsymbol{x}^*$ is a k-sparse vector, we have $|supp(\boldsymbol{b})| > (1-2\epsilon)d|S^*|$, where $S^* = supp(\boldsymbol{x}^*)$.*

Proof of Lemma 2. Since $|S^*| < k+1$, we have $N(S^*) > (1-\epsilon)d|S^*|$ by the expansion property. Now, $N(S^*)$ can be partitioned into $N_1(S^*)$ and $N_{>1}(S^*)$, where $N_1(S^*)$ are the vertices in $N(S^*)$ with only one neighbor in S^* and $N_{>1}(S^*)$ are the vertices in $N(S^*)$ with at least two neighbors in S^*.

So the number of edges between S^* and $N(S^*)$ is $d|S^*| \geq |N_1(S^*)| + 2|N_{>1}(S^*)|$. Also $|N(S^*)| = |N_1(S^*)| + |N_{>1}(S^*)| > (1-\epsilon)d|S^*|$. Eliminating $|N_{>1}(S^*)|$, we obtain $|N_1(S^*)| > (1-2\epsilon)d|S^*|$. Also, $N_1(S) \subseteq \text{supp}(\boldsymbol{b})$. Hence, $|\text{supp}(\boldsymbol{b})| > (1-2\epsilon)d|S^*|$. □

Proof of Theorem 5. We first prove $S^* \subseteq \widehat{S}$. Let $j \in \text{supp}(\boldsymbol{x}^*)$, then $|N(j) \cup \text{supp}(\boldsymbol{b})| \leq |N(S^* \cup j)| \leq d|S^*|$. Using Lemma 2 with the above inequality we get:
$|N(j) \cap \text{supp}(\boldsymbol{b})| > (1-2\epsilon)d|S^*| - d(|S^*|-1)$.
As $\epsilon < \frac{1}{8k}$, $|N(j) \cap \text{supp}(\boldsymbol{b})| > \frac{3d}{4}$. Hence, Step 4 of Algorithm 3 will add j to $\widehat{S}$ and hence, $S^* \subseteq \widehat{S}$.

We now prove $\widehat{S} \subseteq S^*$. Let $j \notin S^*$, then $|S^* \cup \{j\}| \leq k+1$. Using expansion property,

$$\begin{aligned}(1-\epsilon)d(|S^*|+1) &< |N(S^* \cup \{j\})| \\ &\leq |N(S^*)| + |N(j)| - |N(j) \cap N(S^*)| \\ &\leq d|S^*| + d - |N(j) \cap N(S^*)| \end{aligned}$$
$$\Rightarrow |N(j) \cap N(S^*)| < \epsilon d(|S^*|+1) \leq \epsilon d(k+1) < \frac{d}{4}.$$

As $\text{supp}(\boldsymbol{b}) \subseteq N(S^*)$, $|N(j) \cap \text{supp}(\boldsymbol{b})| < \frac{d}{4}$. Hence, Step 4 of Algorithm 3 will *not* add j to $\widehat{S}$. Hence, $\widehat{S} \subseteq S^*$. □

C. Proof of the Divide and Conquer approach

Proof of Theorem 6. Let $r = \log k$, $\boldsymbol{z} = P\boldsymbol{x}^*$ and $\boldsymbol{z}_\ell = \boldsymbol{z}((\ell-1)\frac{m}{k}, \cdots, \ell\frac{m}{k} - 1)$ i.e. the ℓ^{th} block of $\boldsymbol{z}$. Now,

$$\Pr[||\boldsymbol{z}_\ell||_0 > r] \leq \binom{k}{r}\frac{1}{k^r} \leq \left(\frac{e}{r}\right)^r$$

where the second inequality follows from Stirling's approximation. By union bound, we have

$$\Pr[\exists \ell : ||\boldsymbol{z}_\ell||_0 > r] \leq k\left(\frac{e}{r}\right)^r = e^{-\Omega(\log k)}.$$

So $\|\boldsymbol{z}_\ell\|_0$, $\forall \ell$ is at most $\mathcal{O}(\log k)$ with probability at least $1 - e^{-\Omega(\log k)}$. Theorem now follows using Theorem 3. □

D. GraDeS

This section is almost entirely from (Garg & Khandekar, 2009), presented here for the sake of completeness. Before we present the GraDeS algorithm, we have the following definition:

Definition 5. *Let $H_k : \mathbb{R}^n \to \mathbb{R}^n$ be a function that sets all but the k largest coordinates in absolute value to zero. More precisely, for $\boldsymbol{x} \in \mathbb{R}^n$, let π be a permutation of $[n]$ such that $|x_{\pi(1)}| \geq |x_{\pi(2)}| \geq \cdots \geq |x_{\pi(n)}|$. Then the vector $H_k(\boldsymbol{x})$ is a vector $\widehat{\boldsymbol{x}}$ where $\widehat{x}_{\pi(i)} = x_{\pi(i)}$ for $i \leq k$ and $\widehat{x}_{\pi(i)} = 0$ for $i \geq k+1$.*

Algorithm 9 GraDeS (Garg & Khandekar, 2009)

input $\widehat{\boldsymbol{z}}, A_1, \gamma$ and ϵ
Initialize $\widehat{\boldsymbol{x}} \leftarrow 0$
while $\|\widehat{\boldsymbol{z}} - A_1\widehat{\boldsymbol{x}}\|^2 > \epsilon$ **do**
  $\widehat{\boldsymbol{x}} \leftarrow H_k\left(\widehat{\boldsymbol{x}} + \frac{1}{\gamma}A_1^T(\widehat{\boldsymbol{z}} - A_1\widehat{\boldsymbol{x}})\right)$
end while
output $\widehat{\boldsymbol{x}}$

The following theorem which shows the correctness of Algorithm 9 is a restatement of Theorem 2.3 from (Garg & Khandekar, 2009).

Theorem 9. *Suppose $\boldsymbol{x}^*$ is a k-sparse vector satisfying $\widehat{\boldsymbol{z}} = A_1\boldsymbol{x}^* + \boldsymbol{e}$ for an error vector $\boldsymbol{e} \in \mathbb{R}^{m'}$ and the isometry constant of the matrix A_1 satisfies $\delta_{2k} < \frac{1}{3}$. There exists a constant $D > 0$ that depends only on δ_{2k}, such that Algorithm 9 with $\gamma = 1 + \delta_{2k}$, computes a k-sparse vector $\widehat{\boldsymbol{x}} \in \mathbb{R}^n$ satisfying $\|\boldsymbol{x}^* - \widehat{\boldsymbol{x}}\| \leq D\|\boldsymbol{e}\|$ in at most*

$$\left\lceil \frac{1}{\log\left(\frac{1-\delta_{2k}}{4\delta_{2k}}\right)} \cdot \log\left(\frac{\|\widehat{\boldsymbol{z}}\|^2}{\|\boldsymbol{e}\|^2}\right)\right\rceil$$

iterations. Moreover, for $\delta_{2k} < \frac{1}{6}$, we can choose the constant D to be 6.

E. Recovery using Gaussian Measurements

Here we state a theorem from (Jacques et al., 2011) which guarantees that all unit vectors which agree with the 1-bit measurements obtained from a random Gaussian matrix must be very close to each other.

Theorem 10 (Theorem 2 of (Jacques et al., 2011))**.** *Let $A \in \mathbb{R}^{m\times n}$ be a matrix generated as $A \sim \mathcal{N}^{m\times n}(0,1)$. Fix $0<\eta\le 1$ and $\epsilon>0$. If the number of measurements(m) satisfy:*

$$m > \frac{8}{\epsilon} k \log(\frac{16n}{\epsilon\eta}),$$

*then with probability $1-\eta$, **for all** k-sparse vectors $\boldsymbol{x}$ and $\boldsymbol{y}$:*

$$sign(A\boldsymbol{x}) = sign(A\boldsymbol{y}) \Rightarrow \left|\left|\frac{\boldsymbol{x}}{||\boldsymbol{x}||_2} - \frac{\boldsymbol{y}}{||\boldsymbol{y}||_2}\right|\right|_2 \le \epsilon.$$

F. Proof of the Two-stage algorithm (Algorithm 6)

Here we prove Theorem 7 which is a proof of correctness of Two-stage algorithm (Algorithm 6).

Proof of Theorem 7. We prove the theorem by analyzing both the stages of our algorithm.

Stage 1: Let $\boldsymbol{z}^* = A_1\boldsymbol{x}^*$. As $\boldsymbol{b} = \text{sign}(A_2\boldsymbol{z}^*)$, $(\mathbf{a}_2^{(i)}, b_i), \forall i$ are linearly separable and hence using linear programming, we can find a vector $\widehat{\boldsymbol{z}}$ consistent with the measurements $\boldsymbol{b}$ i.e. $\boldsymbol{b} = \text{sign}(A_2\widehat{\boldsymbol{z}})$. Using Theorem 10,

$$\left|\left|\frac{\boldsymbol{z}^*}{||\boldsymbol{z}^*||_2} - \frac{\widehat{\boldsymbol{z}}}{||\widehat{\boldsymbol{z}}||_2}\right|\right|_2 < \epsilon. \tag{6}$$

Stage 2: In stage 2 of Algorithm 6, we run GradeS with inputs $\frac{\widehat{\boldsymbol{z}}}{\|\widehat{\boldsymbol{z}}\|_2}$ and A_1. Now, using (6):

$$\frac{\widehat{\boldsymbol{z}}}{\|\widehat{\boldsymbol{z}}\|_2} = A_1\frac{\boldsymbol{x}^*}{\|A_1\boldsymbol{x}^*\|_2} + \boldsymbol{\eta},$$

where $\|\boldsymbol{\eta}\|_2 \le \epsilon$. Also, since A_1 satisfies RIP with $\delta_{2k} < 1/6$, using the recovery result of GradeS (Theorem 9, Appendix D), the recovered vector $\hat{\boldsymbol{x}}$ satisfies:

$$\left|\left|\hat{\boldsymbol{x}} - \frac{\boldsymbol{x}^*}{\|A_1\boldsymbol{x}^*\|_2}\right|\right|_2 \le 6\epsilon.$$

That is,

$$\|\hat{\boldsymbol{x}}\|_2^2 + \frac{\|\boldsymbol{x}^*\|_2^2}{\|A_1\boldsymbol{x}^*\|_2^2} - 2\frac{\hat{\boldsymbol{x}}^T\boldsymbol{x}^*}{\|A_1\boldsymbol{x}^*\|_2} \le 36\epsilon^2,$$

$$\frac{\|\hat{\boldsymbol{x}}\|_2\|A_1\boldsymbol{x}^*\|_2}{\|\boldsymbol{x}^*\|_2} + \frac{\|\boldsymbol{x}^*\|_2}{\|\hat{\boldsymbol{x}}\|_2\|A_1\boldsymbol{x}^*\|_2} - 2\frac{\hat{\boldsymbol{x}}^T\boldsymbol{x}^*}{\|\hat{\boldsymbol{x}}\|_2\|\boldsymbol{x}^*\|_2} \le 36\epsilon^2\frac{\|A_1\boldsymbol{x}^*\|_2}{\|\hat{\boldsymbol{x}}\|_2\|\boldsymbol{x}^*\|_2}.$$

Using the fact that $t + 1/t \ge 2$ and using RIP,

$$2 - 2\frac{\hat{\boldsymbol{x}}^T\boldsymbol{x}^*}{\|\hat{\boldsymbol{x}}\|_2\|\boldsymbol{x}^*\|_2} \le 36\epsilon^2(1+\delta_{2k})\frac{1}{\|\hat{\boldsymbol{x}}\|_2}.$$

Also, $\|\hat{\boldsymbol{x}}\|_2 \ge \|\frac{\boldsymbol{x}^*}{\|A_1\boldsymbol{x}^*\|_2}\|_2 - 6\epsilon \ge \frac{1}{1+\delta_{2k}} - 6\epsilon$. So we have

$$\left|\left|\frac{\boldsymbol{x}^*}{\|\boldsymbol{x}^*\|_2} - \frac{\hat{\boldsymbol{x}}}{\|\hat{\boldsymbol{x}}\|_2}\right|\right|_2^2 < 36\left(\frac{1+\delta_{2k}}{\left(\frac{1}{1+\delta_{2k}} - \epsilon\right)}\right)\epsilon^2$$

$$\Rightarrow \left|\left|\frac{\boldsymbol{x}^*}{\|\boldsymbol{x}^*\|_2} - \frac{\hat{\boldsymbol{x}}}{\|\hat{\boldsymbol{x}}\|_2}\right|\right|_2 < 20\epsilon$$

for $\epsilon < \frac{1}{4}$. □

G. Lower Bound on Reconstruction Error

The following is a lower bound on the reconstruction error of any approximate recovery algorithm from (Jacques et al., 2011).

Theorem 11 (Theorem 1 of (Jacques et al., 2011))**.** *Let $||\boldsymbol{x}^*||_0 \le k, ||\boldsymbol{x}^*||_2 = 1, \boldsymbol{b} = sign(A\boldsymbol{x}^*), A \in \mathbb{R}^{m\times n}$ and let $\hat{\boldsymbol{x}} = \Delta^{1bit}(\boldsymbol{b}, A, k)$ be the unit vector reconstructed by some recovery algorithm Δ^{1bit} based on $\boldsymbol{b}, A, k$. Then the worst case reconstruction error $\sup_{\boldsymbol{x}^*} ||\hat{\boldsymbol{x}} - \boldsymbol{x}^*||_2 \ge \frac{k}{em}$.*