Supplementary: Robust Recovery of Joint Sparse Signals via Simultaneous Orthogonal Matching Pursuit

Yuxuan Zhang and Jian Wang

School of Data Science, Fudan University, Shanghai 200433, China

This supplementary is dedicated to the proofs for Propositions 1 and 2 and Theorems 1 and 2 in our main paper.

A. Preliminary

We review some useful notations. Let $\Omega := \{1, \cdots, n\}$ and $S :=$ rsupp$(\mathbf{X})$ denote the row support of $\mathbf{X}$. Define $\mathbf{X}^S$ (or $\mathbf{X}_S$) as a sub-matrix of $\mathbf{X}$ with rows (or columns) indexed by S. For $T \subset \Omega$, $S \backslash T$ denotes the set that contains elements in S but not in T. $|S|$ is the cardinality of S. If $\mathbf{\Phi}_T$ has full column rank, $\mathbf{\Phi}_T^\dagger = (\mathbf{\Phi}_T' \mathbf{\Phi}_T)^{-1} \mathbf{\Phi}_T$ is the Moore-Penrose pseudo-inverse of $\mathbf{\Phi}_T$. $\mathcal{R}(\mathbf{\Phi})$ is the vector space spanned by the columns of $\mathbf{\Phi}$ and $\mathbf{P}_T = \mathbf{\Phi}_T \mathbf{\Phi}_T^\dagger$ represents the orthogonal projection onto $\mathcal{R}(\mathbf{\Phi}_T)$.

We also introduce two lemmas, which are useful for proving Proposition 1. The first lemma formulates the projection operator $\mathbf{P}_{\mathcal{R}(\boldsymbol{\phi})} := \boldsymbol{\phi}(\boldsymbol{\phi}^T\boldsymbol{\phi})^{-1}\boldsymbol{\phi}^T$ as a scaling transformation along the direction of vector $\boldsymbol{\phi}$.

Lemma 1. *For any vector $\boldsymbol{\phi} \in \mathbb{R}^m$ and $\mathbf{x} \in \mathbb{R}^m$, the projection operator $\mathbf{P}_{\mathcal{R}(\boldsymbol{\phi})}$ satisfies*

$$\|\mathbf{P}_{\mathcal{R}(\boldsymbol{\phi})}\mathbf{x}\|_2 = \frac{1}{\|\boldsymbol{\phi}\|_2}\|\boldsymbol{\phi}^T\mathbf{x}\|_2. \tag{1}$$

Proof. Since $\boldsymbol{\phi}^T\boldsymbol{\phi}$ is a real number, we have

$$\begin{aligned}\|\mathbf{P}_{\mathcal{R}(\boldsymbol{\phi})}\mathbf{x}\|_2^2 &= \|\boldsymbol{\phi}(\boldsymbol{\phi}^T\boldsymbol{\phi})^{-1}\boldsymbol{\phi}^T\mathbf{x}\|_2^2 \\ &= \frac{1}{(\boldsymbol{\phi}^T\boldsymbol{\phi})^2}\|\boldsymbol{\phi}\boldsymbol{\phi}^T\mathbf{x}\|_2^2 \\ &= \frac{1}{(\boldsymbol{\phi}^T\boldsymbol{\phi})^2}\mathbf{x}^T\boldsymbol{\phi}\boldsymbol{\phi}^T\boldsymbol{\phi}\boldsymbol{\phi}^T\mathbf{x} \\ &= \frac{1}{(\boldsymbol{\phi}^T\boldsymbol{\phi})}\mathbf{x}^T\boldsymbol{\phi}\boldsymbol{\phi}^T\mathbf{x} \\ &= \frac{1}{\|\boldsymbol{\phi}\|_2}\|\boldsymbol{\phi}^T\mathbf{x}\|_2,\end{aligned}$$

which is the desired result. □

The second lemma explores the relationship between the inner product of two matrices and their norms.

Lemma 2. *Let $\mathbf{U} \in \mathbb{R}^{m\times n}$ and $\mathbf{V} \in \mathbb{R}^{m\times n}$ be two matrices supported on rsupp$(\mathbf{U})$ and rsupp$(\mathbf{V})$, respectively. Then,*

$$\langle \mathbf{U}, \mathbf{V} \rangle_F \le \max\{1, |S|^{1/2}\}\|\mathbf{U}^S\|_F|\mathbf{V}^j|_2,$$

where $S :=$ rsupp$(\mathbf{U}) \cup$ rsupp$(\mathbf{V})$ and j is an index corresponding to the row of $\mathbf{V}$ that has the maximum ℓ_2-norm.

Proof. From the definition of S, it is easy to see that if $i \notin S$, $\mathbf{U}^i\mathbf{V}^i = 0$. By employing Cauthy-Schwarz inequality, we have

$$\langle \mathbf{U}, \mathbf{V} \rangle_F = \left\langle \mathbf{U}^S, \mathbf{V}^S \right\rangle_F \le \|\mathbf{U}^S\|_F\|\mathbf{V}^S\|_F. \tag{2}$$

Since j is an index corresponding to the row of $\mathbf{V}$ that has the maximum ℓ_2-norm, we can estimate $\|\mathbf{V}^S\|_F$ as

$$\|\mathbf{V}^S\|_F^2 = \left(\sum_{i\in S}|\mathbf{V}^i|_2\right)^{1/2} \le |S|^{1/2}|\mathbf{V}^j|_2. \tag{3}$$

Combining (2) with (3) yields the desired result. □

B. Proof of Proposition 1

Proposition 1. *Suppose that there are $\lambda^k = |\Lambda^k|$ remaining support indices after k iterations of SOMP. Let $j \ge k$ be an arbitrary integer. Then, for any $1 \le \tau \le \lfloor \log_2 \lambda^k \rfloor + 1$, the following inequality holds.*

$$\begin{aligned}\left\|\mathbf{R}^j\right\|_F^2 - \left\|\mathbf{R}^{j+1}\right\|_F^2 \ge{}& \frac{\left(1-\delta_{|\Lambda_\tau^k \cup S^j|}\right)^2}{\max\{1, |\Lambda_\tau^k \backslash S^j|\}} \\ &\times \left(\left\|\mathbf{R}^j\right\|_F^2 - \left\|\mathbf{\Phi}_{\Lambda^k\backslash\Lambda_\tau^k}\mathbf{X}^{\Lambda^k\backslash\Lambda_\tau^k} + \mathbf{W}\right\|_F^2\right).\end{aligned} \tag{4}$$

This proposition offers a lower bound on the energy reduction, indicating that SOMP makes a non-trivial progress in each iteration.

Proof. Our proof consists of two steps. First, we give an explicit representation for the residual reduction $\|\mathbf{R}^j\|_F^2 - \|\mathbf{R}^{j+1}\|_F^2$ in one iteration of SOMP, expressed in terms of orthogonal projection:

$$\|\mathbf{R}^j\|_F^2 - \|\mathbf{R}^{j+1}\|_F^2 = \|\mathbf{P}_{\mathcal{R}(\mathbf{P}_{S^j}^\perp \boldsymbol{\phi}_{s^{j+1}})}\mathbf{R}^j\|_F^2. \tag{5}$$

Then, we use (5) to derive a lower bound for $\|\mathbf{R}^j\|_F^2 - \|\mathbf{R}^{j+1}\|_F^2$.

1. Proof of Eq. (5):

Recall that the residual of SOMP is the orthogonal projection of measurement vectors $\mathbf{Y}$ onto the orthogonal complement space of $\mathcal{R}(\mathbf{\Phi}_{S^j})$ (i.e, $\mathbf{R}^j = \mathbf{P}_{S^j}^\perp\mathbf{Y}$). Then, we have

$$\mathbf{R}^{j+1} = \mathbf{P}_{S^{j+1}}^\perp\mathbf{Y} = \mathbf{P}_{S^{j+1}}^\perp(\mathbf{R}^j + \mathbf{\Phi}\hat{\mathbf{X}}^j) \overset{(a)}{=} \mathbf{P}_{S^{j+1}}^\perp\mathbf{R}^j,$$

where (a) is from $\mathbf{\Phi}\hat{\mathbf{X}}^j = \mathbf{\Phi}_{S^j}\hat{\mathbf{X}}^{S^j} \in \mathcal{R}(\mathbf{\Phi}_{S^j})$ and $S^j \subseteq S^{j+1}$. Hence, $\left\langle \mathbf{R}^{j+1}, \mathbf{R}^j - \mathbf{R}^{j+1} \right\rangle_F = 0$ and

$$\mathbf{R}^j - \mathbf{R}^{j+1} = \mathbf{R}^j - \mathbf{P}_{S^{j+1}}^\perp\mathbf{R}^j = \mathbf{P}_{S^{j+1}}\mathbf{R}^j.$$

The residual reduction can be rewritten as

$$\begin{aligned}&\|\mathbf{R}^j\|_F^2 - \|\mathbf{R}^{j+1}\|_F^2 \\ ={}& \left\langle \mathbf{R}^j - \mathbf{R}^{j+1}, \mathbf{R}^j - \mathbf{R}^{j+1} \right\rangle_F + 2\left\langle \mathbf{R}^{j+1}, \mathbf{R}^j - \mathbf{R}^{j+1} \right\rangle_F \\ ={}& \|\mathbf{R}^j - \mathbf{R}^{j+1}\|_F^2\end{aligned}$$

$$
\begin{aligned}
&= \|\mathbf{P}_{S^{j+1}}\mathbf{R}^j\|_F^2 \\
&= \|\mathbf{\Phi}_{S^{j+1}}\mathbf{\Phi}_{S^{j+1}}^{\dagger}\mathbf{R}^j\|_F^2 \\
&= \|\mathbf{\Phi}_{S^{j+1}}(\mathbf{\Phi}_{S^{j+1}}^T\mathbf{\Phi}_{S^{j+1}})^{-1}\mathbf{\Phi}_{S^{j+1}}^T\mathbf{R}^j\|_F^2. \qquad (6)
\end{aligned}
$$

Note that $S^{j+1} = S^j \cup s^{j+1}$, by expressing $\mathbf{\Phi}_{S^{j+1}}$ in the form of block matrix as $\mathbf{\Phi}_{S^{j+1}} = [\mathbf{\Phi}_{S^j}\ \ \boldsymbol{\phi}_{s^{j+1}}]$, we have

$$
\mathbf{\Phi}_{S^{j+1}}^T\mathbf{\Phi}_{S^{j+1}} = \begin{bmatrix} \mathbf{\Phi}_{S^j}^T\mathbf{\Phi}_{S^j} & \mathbf{\Phi}_{S^j}^T\boldsymbol{\phi}_{s^{j+1}} \\ \boldsymbol{\phi}_{s^{j+1}}^T\mathbf{\Phi}_{S^j} & \boldsymbol{\phi}_{s^{j+1}}^T\boldsymbol{\phi}_{s^{j+1}} \end{bmatrix}.
$$

Then, by the inverse formula for block matrix, we have

$$
(\mathbf{\Phi}_{S^{j+1}}^T\mathbf{\Phi}_{S^{j+1}})^{-1} = \begin{bmatrix} \mathbf{B} & \mathbf{C} \\ \mathbf{D} & \mathbf{E} \end{bmatrix},
$$

where

$$
\begin{aligned}
\mathbf{B} &:= (\mathbf{\Phi}_{S^j}^T(\mathbf{I} - \boldsymbol{\phi}_{s^{j+1}}\boldsymbol{\phi}_{s^{j+1}}^{\dagger})\mathbf{\Phi}_{S^j})^{-1}, \\
\mathbf{E} &:= (\boldsymbol{\phi}_{s^{j+1}}^T(\mathbf{I} - \mathbf{\Phi}_{S^j}\mathbf{\Phi}_{S^j}^{\dagger})\boldsymbol{\phi}_{s^{j+1}})^{-1}, \\
\mathbf{C} &:= -(\mathbf{\Phi}_{S^j}^T\mathbf{\Phi}_{S^j})^{-1}\mathbf{\Phi}_{S^j}^T\boldsymbol{\phi}_{s^{j+1}}\mathbf{E}, \\
\mathbf{D} &:= -(\boldsymbol{\phi}_{s^{j+1}}^T\boldsymbol{\phi}_{s^{j+1}})^{-1}\boldsymbol{\phi}_{s^{j+1}}^T\mathbf{\Phi}_{S^j}\mathbf{B}.
\end{aligned}
$$

Hence, we can simplify the expression of (6) as

$$
\begin{aligned}
&\|\mathbf{\Phi}_{S^{j+1}}(\mathbf{\Phi}_{S^{j+1}}^T\mathbf{\Phi}_{S^{j+1}})^{-1}\mathbf{\Phi}_{S^{j+1}}^T\mathbf{R}^j\|_F^2. \\
\overset{(a)}{=}\ &\|(\mathbf{\Phi}_{S^{j+1}}(\mathbf{\Phi}_{S^{j+1}}^T\mathbf{\Phi}_{S^{j+1}})^{-1})_{s^{j+1}}\mathbf{\Phi}_{S^{j+1}}^T\mathbf{R}^j\|_F^2 \\
=\ &\|(\mathbf{\Phi}_{S^j}\mathbf{C} + \boldsymbol{\phi}_{s^{j+1}}\mathbf{E})\mathbf{\Phi}_{S^{j+1}}^T\mathbf{R}^j\|_F^2 \\
=\ &\|(\mathbf{I} - \mathbf{\Phi}_{S^j}\mathbf{\Phi}_{S^j}^{\dagger})\boldsymbol{\phi}_{s^{j+1}}\mathbf{E}\mathbf{\Phi}_{S^{j+1}}^T\mathbf{R}^j\|_F^2 \\
=\ &\|\mathbf{P}_{\mathbf{\Phi}_{S^j}}^{\perp}\boldsymbol{\phi}_{s^{j+1}}(\boldsymbol{\phi}_{s^{j+1}}^T\mathbf{P}_{\mathbf{\Phi}_{S^j}}^{\perp}\boldsymbol{\phi}_{s^{j+1}})^{-1}\mathbf{\Phi}_{S^{j+1}}^T\mathbf{R}^j\|_F^2 \\
=\ &\|\mathbf{P}_{\mathbf{\Phi}_{S^j}}^{\perp}\boldsymbol{\phi}_{s^{j+1}}(\boldsymbol{\phi}_{s^{j+1}}^T\mathbf{P}_{\mathbf{\Phi}_{S^j}}^{\perp}\boldsymbol{\phi}_{s^{j+1}})^{-1} \\
&(\mathbf{\Phi}_{S^{j+1}}^T\mathbf{P}_{\mathbf{\Phi}_{S^j}}^{\perp}\mathbf{R}^j)^{s^{j+1}}\|_F^2 \\
=\ &\|\mathbf{P}_{\mathbf{\Phi}_{S^j}}^{\perp}\boldsymbol{\phi}_{s^{j+1}}(\boldsymbol{\phi}_{s^{j+1}}^T\mathbf{P}_{\mathbf{\Phi}_{S^j}}^{\perp}\boldsymbol{\phi}_{s^{j+1}})^{-1}\boldsymbol{\phi}_{s^{j+1}}^T\mathbf{P}_{\mathbf{\Phi}_{S^j}}^{\perp}\mathbf{R}^j\|_F^2 \\
=\ &\|\mathbf{P}_{\mathcal{R}(\mathbf{P}_{\mathbf{\Phi}_{S^j}}^{\perp}\boldsymbol{\phi}_{s^{j+1}})}\mathbf{R}^j\|_F^2,
\end{aligned}
$$

where (a) is due to the fact that $\text{rsupp}(\mathbf{\Phi}_{S^{j+1}}^T\mathbf{R}^j) = s^{j+1}$.

2. Lower bound for $\|\mathbf{R}^j\|_F^2 - \|\mathbf{R}^{j+1}\|_F^2$:

From (5), we have

$$
\begin{aligned}
\|\mathbf{R}^j\|_F^2 - \|\mathbf{R}^{j+1}\|_F^2 &= \|\mathbf{P}_{\mathcal{R}(\mathbf{P}_{S^j}^{\perp}\boldsymbol{\phi}_{s^{j+1}})}\mathbf{R}^j\|_F^2 \\
&= \sum_{i=1}^{r}\|\mathbf{P}_{\mathcal{R}(\mathbf{P}_{S^j}^{\perp}\boldsymbol{\phi}_{s^{j+1}})}\mathbf{r}_i^j\|_2^2 \\
&\overset{(a)}{=} \frac{1}{\|\mathbf{P}_{S^j}^{\perp}\boldsymbol{\phi}_{s^{j+1}}\|_2}\sum_{i=1}^{r}\|(\mathbf{r}_i^j)^T\mathbf{P}_{S^j}^{\perp}\boldsymbol{\phi}_{s^{j+1}}\|_2^2 \\
&= \frac{1}{\|\mathbf{P}_{S^j}^{\perp}\boldsymbol{\phi}_{s^{j+1}}\|_2}\|(\mathbf{R}^j)^T\mathbf{P}_{S^j}^{\perp}\boldsymbol{\phi}_{s^{j+1}}\|_2^2 \\
&\overset{(b)}{\geq} \frac{1}{\|\boldsymbol{\phi}_{s^{j+1}}\|_2}\|(\mathbf{R}^j)^T\mathbf{P}_{S^j}^{\perp}\boldsymbol{\phi}_{s^{j+1}}\|_2^2 \\
&\overset{(c)}{=} \max_{i\in\Omega\backslash S^j}\|(\mathbf{R}^j)^T\mathbf{P}_{S^j}^{\perp}\boldsymbol{\phi}_i\|_2^2 \\
&= \max_{i\in\Omega\backslash S^j}\|(\boldsymbol{\phi}_i)^T\mathbf{R}^j\|_2^2, \qquad (7)
\end{aligned}
$$

where (a) is from Lemma 1, (b) is owing to that orthogonal projection cannot increase the norm of the vector, and (c) is based on the selection criterion of SOMP and that $\mathbf{\Phi}$ has ℓ_2-normalized columns. Thus, to build a lower bound for $\|\mathbf{R}^j\|_F^2 - \|\mathbf{R}^{j+1}\|_F^2$, it suffices to analyze $\max_{i\in\Omega\backslash S^j}\|(\boldsymbol{\phi}_i)^T\mathbf{R}^j\|_2^2$.

Note that $\text{rsupp}(\mathbf{\Phi}^T\mathbf{R}^j) = \Omega\backslash S^j$, we define $\mathbf{H} \in \mathbb{R}^{n\times r}$ as $\mathbf{H}^{S\cap S^k\cup\Lambda_\tau^k} = \mathbf{X}^{S\cap S^k\cup\Lambda_\tau^k}$ and $\mathbf{H}^{(S\cap S^k\cup\Lambda_\tau^k)^c} = 0$. Then, from Lemma 2 and also noting that $\text{rsupp}(\mathbf{H}) \cap \text{rsupp}(\mathbf{\Phi}^T\mathbf{R}^j) = \Lambda_\tau^k\backslash S^j$, we have

$$
\begin{aligned}
&\left\langle \mathbf{\Phi}^T\mathbf{R}^j, \mathbf{H}\right\rangle_F \\
&\leq \max\left\{1, |\Lambda_\tau^k\backslash S^j|^{\frac{1}{2}}\right\}\max_{i\in\Omega\backslash S^j}\|\boldsymbol{\phi}_i^T\mathbf{R}^j\|_2\|\mathbf{H}^{\Lambda_\tau^k\backslash S^j}\|_F \\
&\overset{(a)}{\leq} \max\left\{1, |\Lambda_\tau^k\backslash S^j|^{\frac{1}{2}}\right\}\max_{i\in\Omega\backslash S^j}\|\boldsymbol{\phi}_i^T\mathbf{R}^j\|_2\|\mathbf{H}^{\Omega\backslash S^j}\|_F.
\end{aligned}
$$

where (a) is because $\Lambda_\tau^k\backslash S^j \subseteq \Omega\backslash S^j$. Hence, we can derive that

$$
\max_{i\in\Omega\backslash S^j}\|\boldsymbol{\phi}_i^T\mathbf{R}^j\|_2 \geq \frac{\left\langle \mathbf{\Phi}^T\mathbf{R}^j, \mathbf{H}\right\rangle_F}{\max\left\{1, |\Lambda_\tau^k\backslash S^j|^{\frac{1}{2}}\right\}}\|\mathbf{H}^{\Omega\backslash S^j}\|_F. \qquad (8)
$$

Since $\|\mathbf{H}^{\Omega\backslash S^j}\|_F$ and $\max\left\{1, |\Lambda_\tau^k\backslash S^j|^{1/2}\right\}$ are not related with residual, it suffices to provide a lower bound for the inner product $\left\langle \mathbf{\Phi}^T\mathbf{R}^j, \mathbf{H}\right\rangle_F$.

$$
\begin{aligned}
&\left\langle \mathbf{\Phi}^T\mathbf{R}^j, \mathbf{H}\right\rangle_F \\
\overset{(a)}{=}\ &\langle \mathbf{\Phi}^T\mathbf{R}^j, \mathbf{H} - \hat{\mathbf{X}}^j\rangle_F \\
\overset{(b)}{=}\ &\langle \mathbf{R}^j, \mathbf{\Phi}(\mathbf{H} - \hat{\mathbf{X}}^j)\rangle_F \\
\overset{(c)}{=}\ &\frac{1}{2}\left(\|\mathbf{\Phi}(\mathbf{H} - \hat{\mathbf{X}}^j)\|_F^2 + \|\mathbf{R}^j\|_F^2 - \|\mathbf{\Phi}(\mathbf{H} - \hat{\mathbf{X}}^j) - \mathbf{R}^j\|_F^2\right) \\
\overset{(d)}{=}\ &\frac{1}{2}\left(\|\mathbf{\Phi}(\mathbf{H} - \hat{\mathbf{X}}^j)\|_F^2 + \|\mathbf{R}^j\|_F^2 - \|\mathbf{\Phi}(\mathbf{X} - \mathbf{H}) + \mathbf{W}\|_F^2\right) \\
=\ &\frac{1}{2}\left(\|\mathbf{\Phi}(\mathbf{H} - \hat{\mathbf{X}}^j)\|_F^2 + \|\mathbf{R}^j\|_F^2 - \|\mathbf{\Phi}_{\Lambda^k\backslash\Lambda_\tau^k}\mathbf{X}^{\Lambda^k\backslash\Lambda_\tau^k} + \mathbf{W}\|_F^2\right) \\
\overset{(e)}{\geq}\ &\|\mathbf{\Phi}(\mathbf{H} - \hat{\mathbf{X}}^j)\|_F\left(\|\mathbf{R}^j\|_F^2 - \|\mathbf{\Phi}_{\Lambda^k\backslash\Lambda_\tau^k}\mathbf{X}^{\Lambda^k\backslash\Lambda_\tau^k} + \mathbf{W}\|_F^2\right)^{1/2} \\
\overset{(f)}{\geq}\ &\left(1 - \delta_{|\Lambda_\tau^k\cup S^j|}\right)\|\mathbf{H} - \hat{\mathbf{X}}^j\|_F \\
&\times\left(\|\mathbf{R}^j\|_F^2 - \|\mathbf{\Phi}_{\Lambda^k\backslash\Lambda_\tau^k}\mathbf{X}^{\Lambda^k\backslash\Lambda_\tau^k} + \mathbf{W}\|_F^2\right)^{1/2} \\
\geq\ &\left(1 - \delta_{|\Lambda_\tau^k\cup S^j|}\right)\|\mathbf{H}^{\Omega\backslash S^j}\|_F \\
&\times\left(\|\mathbf{R}^j\|_F^2 - \|\mathbf{\Phi}_{\Lambda^k\backslash\Lambda_\tau^k}\mathbf{X}^{\Lambda^k\backslash\Lambda_\tau^k} + \mathbf{W}\|_F^2\right)^{1/2},
\end{aligned}
$$

where (a) comes from $\text{rsupp}(\mathbf{\Phi}^T\mathbf{R}^j) = \Omega\backslash S^j$ and $\text{rsupp}(\mathbf{X}^j) = S^j$, (b) is the property of ad-joint operator, (c) is based on the cosine law, (d) relies on the fact that $\mathbf{R}^j = \mathbf{\Phi}(\mathbf{X} - \hat{\mathbf{X}}^j) + \mathbf{W}$, (e) is the fundamental inequality and (f) uses the RIP of the measurement matrix $\mathbf{\Phi}$. This bound, combined with (8), leads to the following result.

$$
\begin{aligned}
\max_{i\in\Omega\backslash S^j}\|(\boldsymbol{\phi}_i)^T\mathbf{R}^j\|_2 &\geq \frac{1 - \delta_{|\Lambda_\tau^k\cup S^j|}}{\max\{1, |\Lambda_\tau^k\backslash S^j|^{1/2}\}} \\
&\quad\times\left(\|\mathbf{R}^j\|_F^2 - \|\mathbf{\Phi}_{\Lambda^k\backslash\Lambda_\tau^k}\mathbf{X}^{\Lambda^k\backslash\Lambda_\tau^k} + \mathbf{W}\|_F^2\right)^{1/2}. \qquad (9)
\end{aligned}
$$

Combining (7) and (9), we complete the proof. □

C. Proof of Proposition 2

Proposition 2. *Suppose that there are $\lambda^k = |\Lambda^k|$ remaining support indices after k iterations of SOMP. Let $j \geq k$ be an* *arbitrary integer. Then, for any* $1 \leq \tau \leq \lfloor \log_2 \lambda^k \rfloor + 1$ *and any integer* $\Delta_j > 0$*, we have*

$$\begin{aligned}&\|\mathbf{R}^{j+\Delta_j}\|_F^2 - \|\mathbf{\Phi}_{\Lambda^k \backslash \Lambda_\tau^k}\mathbf{X}^{\Lambda^k \backslash \Lambda_\tau^k} + \mathbf{W}\|_F^2 \\ &\geq C_{\tau,j,\Delta_j}\big(\|\mathbf{R}^j\|_F^2 - \|\mathbf{\Phi}_{\Lambda^k \backslash \Lambda_\tau^k}\mathbf{X}^{\Lambda^k \backslash \Lambda_\tau^k} + \mathbf{W}\|_F^2\big), \qquad (10)\end{aligned}$$

where $C_{\tau,j,\Delta_j} = \exp\left(-\frac{\Delta_j(1-\delta_{|\Lambda_\tau^k \cup S^{j+\Delta_j}|})^2}{\max\{1,|\Lambda_\tau^k \backslash S^j|\}}\right)$ *is a constant.*

This proposition extends Proposition 1 to characterize the residual reduction after multiple iterations of SOMP.

Proof. For any $j_0 \in \{j, \cdots, j+\Delta j - 1\}$, we apply Proposition 1 to get

$$\begin{aligned}&\|\mathbf{R}^{j_0}\|_F^2 - \|\mathbf{R}^{j_0+1}\|_F^2 \\ \geq& \frac{\left(1-\delta_{|\Lambda_\tau^k \cup S^{j_0}|}\right)^2}{\max\{1,|\Lambda_\tau^k \backslash S^{j_0}|\}}\left(\|\mathbf{R}^{j_0}\|_F^2 - \|\mathbf{\Phi}_{\Lambda^k \backslash \Lambda_\tau^k}\mathbf{X}^{\Lambda^k \backslash \Lambda_\tau^k} + \mathbf{W}\|_F^2\right) \\ \overset{(a)}{\geq}& \left(1-\exp\left(-\frac{\left(1-\delta_{|\Lambda_\tau^k \cup S^{j_0}|}\right)^2}{\max\{1,|\Lambda_\tau^k \backslash S^{j_0}|\}}\right)\right) \\ &\times \left(\|\mathbf{R}^{j_0}\|_F^2 - \|\mathbf{\Phi}_{\Lambda^k \backslash \Lambda_\tau^k}\mathbf{X}^{\Lambda^k \backslash \Lambda_\tau^k} + \mathbf{W}\|_F^2\right) \\ \overset{(b)}{\geq}& \left(1-\exp\left(-\frac{\left(1-\delta_{|\Lambda_\tau^k \cup S^{j+\Delta j-1}|}\right)^2}{\max\{1,|\Lambda_\tau^k \backslash S^{j}|\}}\right)\right) \\ &\times \left(\|\mathbf{R}^{j_0}\|_F^2 - \|\mathbf{\Phi}_{\Lambda^k \backslash \Lambda_\tau^k}\mathbf{X}^{\Lambda^k \backslash \Lambda_\tau^k} + \mathbf{W}\|_F^2\right),\end{aligned}$$

where (a) comes from the inequality that $e^x > 1+x$ and (b) uses the monotonicity of RIC with $j \leq j_0 \leq j+\Delta j - 1$.

Through some transformation, we further have

$$\begin{aligned}&\|\mathbf{R}^{j_0+1}\|_F^2 - \|\mathbf{\Phi}_{\Lambda^k \backslash \Lambda_\tau^k}\mathbf{X}^{\Lambda^k \backslash \Lambda_\tau^k} + \mathbf{W}\|_F^2 \\ &\leq \exp\left(-\frac{\left(1-\delta_{|\Lambda_\tau^k \cup S^{j+\Delta j-1}|}\right)^2}{\max\{1,|\Lambda_\tau^k \backslash S^{j}|\}}\right) \\ &\qquad \times \left(\|\mathbf{R}^{j_0}\|_F^2 - \|\mathbf{\Phi}_{\Lambda^k \backslash \Lambda_\tau^k}\mathbf{X}^{\Lambda^k \backslash \Lambda_\tau^k} + \mathbf{W}\|_F^2\right).\end{aligned}$$

By combining all these inequalities from j to $j+\Delta j-1$, we obtain the conclusion. □

D. Proof of Theorem 1

Theorem 1. *Suppose that there are* $\lambda^k = |\Lambda^k|$ *remaining support indices after* k *iterations of SOMP. Also, suppose that the measurement matrix* $\mathbf{\Phi}$ *has* ℓ_2*-normalized columns and satisfies the RIP of order* $\max\{p_1, p_2\}$*, where* $p_1 = \lambda^k + k + 1$ *and* $p_2 = k + \lceil (c+1)\lambda^k \rceil$*. Then, the residual of SOMP obeys*

$$\left\|\mathbf{R}^{k+\lceil c\lambda^k \rceil}\right\|_F \leq C_0 \|\mathbf{W}\|_F \qquad (11)$$

if $c > \max\{c', c''\}$ *for some constant* $t > 0$*, where*

$$c' = \frac{-4}{(1-\delta_{p_2})^2}\ln\left(\frac{1}{2} - \frac{1}{2(1-\delta_{p_1})^2} + \frac{1}{2(1-\delta_{p_1})(1+t)(1+\delta_{\lambda^k})}\right),$$

$$c'' = \frac{-4}{(1-\delta_{p_2})^2}\ln\left(\frac{1}{2} - \frac{1}{2}\sqrt{1-\frac{1-\delta_{p_2}}{(1+t)(1+\delta_{\lambda^k})}}\right),$$

and C_0 *is an absolute constant.*

The proof of Theorem 1 is based on mathematical induction in the number of the remaining support indices λ^k. We first consider $\lambda^k = 0$. Since SOMP has already chosen all support indices, the conclusion is trivial. Then, assume that the statement holds up to an integer N, (i.e., $\lambda^k < N$). Under this assumption, we will validate the conclusion for $\lambda^k = N$.

Our proof for this case is divided into two cases (i.e., $L = 1$ and $L \geq 2$), depending on how many support indices are still needed to be selected. In particular, according to Definition 1 in our main paper, $L = 1$ indicates that the energy of the "remaining" signal $\mathbf{X}^{\Lambda^k}$ is well concentrated on the subset Λ_1^k, which contains only one index. Whereas, $L \geq 2$ indicates that the energy of the "remaining" signal $\mathbf{X}^{\Lambda^k}$ is dispersed over multiple subsets of Λ^k.

- Case 1: L **= 1**

In this case, we need to show that SOMP will select the support index in Λ_1^k in the $(k+1)$th iteration. Instead of proving this result directly, we provide a sufficient condition using energy comparison as

$$\|\mathbf{X}^{\Lambda^{k+1}}\|_F < \|\mathbf{X}^{\Lambda^k}\|_F. \qquad (12)$$

To realize this goal, we employ the well-known nest approximation (see, e.g., [1]). Concretely speaking, we provide an upper bound and a lower bound for the left- and right-hand side of (12), respectively.

- **Upper bound for** $\|\mathbf{X}^{\Lambda^{k+1}}\|_F$
  Since $\mathbf{R}^{k+1} = \mathbf{\Phi}(\mathbf{X} - \hat{\mathbf{X}}^{k+1}) + \mathbf{W}$, we have

$$\begin{aligned}\|\mathbf{R}^{k+1}\|_F =& \|\mathbf{\Phi}(\mathbf{X} - \hat{\mathbf{X}}^{k+1}) + \mathbf{W}\|_F \\ \overset{(a)}{\geq}& \|\mathbf{\Phi}(\mathbf{X} - \hat{\mathbf{X}}^{k+1})\|_F - \|\mathbf{W}\|_F \\ \geq& \left(1-\delta_{|S^{k+1}\cup S|}\right)^{1/2}\|\mathbf{X} - \hat{\mathbf{X}}^{k+1}\|_F - \|\mathbf{W}\|_F \\ \overset{(b)}{\geq}& \left(1-\delta_{|S^{k+1}\cup S|}\right)^{1/2}\|\mathbf{X}^{\Lambda^{k+1}}\|_F - \|\mathbf{W}\|_F \\ \overset{(c)}{\geq}& (1-\delta_{p_1})^{1/2}\|\mathbf{X}^{\Lambda^{k+1}}\|_F - \|\mathbf{W}\|_F,\end{aligned}$$

  where (a) comes from the triangle inequality, (b) is because $(\hat{\mathbf{X}}^{k+1})^{\Lambda^{k+1}} = 0$ and (c) relies on the monotonicity of RIC and the fact that $|S^{k+1} \cup S| = |S^{K+1} \cup \Lambda^{k+1}| \leq |S^{k+1}| + |\Lambda^k| = \lambda^k + k + 1 = p_1$. Therefore, the upper bound is

$$\|\mathbf{X}^{\Lambda^{k+1}}\|_F \leq \frac{\|\mathbf{R}^{k+1}\|_F + \|\mathbf{W}\|_F}{(1-\delta_{p_1})^{1/2}}. \qquad (13)$$

- **Lower bound for** $\|\mathbf{X}^{\Lambda^k}\|_F$
  From the one-step residual analysis (Proposition 1), we have

$$\begin{aligned}&\|\mathbf{R}^k\|_F^2 - \|\mathbf{R}^{k+1}\|_F^2 \\ \geq& \frac{(1-\delta_{|\Lambda_1^k \cup S^k|})^2}{\max\{1,|\Lambda_1^k \backslash S^k|\}}(\|\mathbf{R}^k\|_F^2 - \|\mathbf{\Phi}_{\Lambda^k \backslash \Lambda_1^k}\mathbf{X}^{\Lambda^k \backslash \Lambda_1^k} + \mathbf{W}\|_F^2) \\ \overset{(a)}{=}& (1-\delta_{|\Lambda_1^k \cup S^k|})^2(\|\mathbf{R}^k\|_F^2 - \|\mathbf{\Phi}_{\Lambda^k \backslash \Lambda_1^k}\mathbf{X}^{\Lambda^k \backslash \Lambda_1^k} + \mathbf{W}\|_F^2) \\ \overset{(b)}{\geq}& (1-\delta_{p_1})^2(\|\mathbf{R}^k\|_F^2 - \|\mathbf{\Phi}_{\Lambda^k \backslash \Lambda_1^k}\mathbf{X}^{\Lambda^k \backslash \Lambda_1^k} + \mathbf{W}\|_F^2), \qquad (14)\end{aligned}$$

  where (a) is because $|\Lambda_1^k \backslash S^k| = 1$ and (b) comes from the fact that $|\Lambda_1^k \cup S^k| \leq |\Lambda^k| + |S^k| \leq \lambda^k + k + 1 = p_1$.

Hence, it is necessary to conduct estimations for $\|\mathbf{R}^k\|_F$ and $\|\mathbf{\Phi}_{\Lambda^k\setminus\Lambda_1^k}\mathbf{X}^{\Lambda^k\setminus\Lambda_1^k}+\mathbf{W}\|_F$. We derive that

$$\begin{aligned}\|\mathbf{R}^k\|_F^2 =&\|\mathbf{P}_{S^k}^{\perp}(\mathbf{\Phi X}+\mathbf{W})\|_F^2\\ &\overset{(a)}{\leq}(1+t)\|\mathbf{P}_{S^k}^{\perp}\mathbf{\Phi X}\|_F^2+\big(1+\frac{1}{t}\big)\|\mathbf{W}\|_F^2\\ &\overset{(b)}{\leq}(1+t)(1+\delta_{\lambda^k})\|\mathbf{X}^{\Lambda^k}\|_F^2+\big(1+\frac{1}{t}\big)\|\mathbf{W}\|_F^2,\end{aligned} \tag{15}$$

where (a) is based on the inequality $(a+b)^2\leq(1+t)a^2+\big(1+\frac{1}{t}\big)b^2$ and (b) comes from the fact that $\mathbf{\Phi}$ satisfies the RIP and $\mathrm{rsupp}(\mathbf{P}_{S^k}^{\perp}\mathbf{\Phi X})=\Lambda^k$.

$$\begin{aligned}&\|\mathbf{\Phi}_{\Lambda^k\setminus\Lambda_1^k}\mathbf{X}^{\Lambda^k\setminus\Lambda_1^k}+\mathbf{W}\|_F^2\\ \leq&(1+t)\|\mathbf{\Phi}_{\Lambda^k\setminus\Lambda_1^k}\mathbf{X}^{\Lambda^k\setminus\Lambda_1^k}\|_F^2+\big(1+\frac{1}{t}\big)\|\mathbf{W}\|_F^2\\ \overset{(a)}{\leq}&(1+t)(1+\delta_{|\Lambda^k\setminus\Lambda_1^k|})\|\mathbf{X}^{\Lambda^k\setminus\Lambda_1^k}\|_F^2+\big(1+\frac{1}{t}\big)\|\mathbf{W}\|_F^2\\ \overset{(b)}{\leq}&(1+t)(1+\delta_{\lambda^k})\frac{1}{\beta}\|\mathbf{X}^{\Lambda^k}\|_F^2+\big(1+\frac{1}{t}\big)\|\mathbf{W}\|_F^2,\end{aligned} \tag{16}$$

where (a) comes from the fact that $\mathbf{\Phi}$ satisfies the RIP and $\mathrm{rsupp}(\mathbf{X}^{\Lambda^k\setminus\Lambda_1^k})=\Lambda^k\setminus\Lambda_1^k$ and (b) depends on the definition of critical point and the monotonicity of RIC. Combining (14), (15) and (16), we obtain

$$\begin{aligned}&\|\mathbf{R}^{k+1}\|_F^2\\ \overset{(14)}{\leq}&\ (1-(1-\delta_{p_1})^2)\|\mathbf{R}^k\|_F^2\\ &+(1-\delta_{p_1})^2\|\mathbf{\Phi}_{\Lambda^k\setminus\Lambda_1^k}\mathbf{X}^{\Lambda^k\setminus\Lambda_1^k}+\mathbf{W}\|_F^2\\ \overset{(14),(15)}{\leq}&\ (1+t)(1+\delta_{\lambda^k})\big((1-(1-\delta_{p_1})^2)+\frac{(1-\delta_{p_1})^2}{\beta}\big)\\ &\times\|\mathbf{X}^{\Lambda^k}\|_F^2+(1+\frac{1}{t})\|\mathbf{W}\|_F^2.\end{aligned}$$

Recall that $\beta=\frac{1}{2}\exp\big(\frac{c}{4}(1-\delta_{p_2})^2\big)$, from the inequality $\sqrt{a^2+b^2}\leq a+b$, it becomes

$$\|\mathbf{R}^{k+1}\|_F^2\leq\mu_0\|\mathbf{X}^{\Lambda^k}\|_F+(1+\frac{1}{t})^{1/2}\|\mathbf{W}\|_F, \tag{17}$$

where $\mu_0=(1+t)^{1/2}(1+\delta_{\lambda^k})^{1/2}\big((1-(1-\delta_{p_1})^2)+\frac{(1-\delta_{p_1})^2}{\beta}\big)^{1/2}$. Hence, we establish a lower bound for $\|\mathbf{X}^{\Lambda^k}\|_F$ as:

$$\|\mathbf{X}^{\Lambda^k}\|_F\geq\frac{1}{\mu_0}(\|\mathbf{R}^{k+1}\|_F^2-\big(1+\frac{1}{t}\big)^{1/2}\|\mathbf{W}\|_F).$$

From (13) and (17), we obtain the final relationship:

$$\|\mathbf{X}^{\Lambda^{k+1}}\|_F\leq\mu_1\|\mathbf{X}^{\Lambda^k}\|_F+\mu_2\|\mathbf{W}\|_F,$$

where $\mu_1=(1-\delta_{p_1})^{-1/2}\mu_0$ and $\mu_2=(1-\delta_{p_1})^{-1/2}\big(1+(1+\frac{1}{t})^{1/2}\big)$.

The first condition $c>c'$ ensures that $\mu_1<1$. Therefore, (12) obviously holds true when $\|\mathbf{W}\|_F<\frac{1-\mu_1}{\mu_2}\|\mathbf{X}^{\Lambda^k}\|_F$.

On the other hand, when $\|\mathbf{W}\|_F\geq\frac{1-\mu_1}{\mu_2}\|\mathbf{X}^{\Lambda^k}\|_F$, we directly prove the theorem as

$$\begin{aligned}&\|\mathbf{R}^{k+\lceil c\lambda^k\rceil}\|_F^2\\ \overset{(a)}{\leq}&(1+t)^{1/2}(1+\delta_{\lambda^k})^{1/2}\|\mathbf{X}^{\Lambda^k}\|_F+\big(1+\frac{1}{t}\big)^{1/2}\|\mathbf{W}\|_F\\ \leq&(1+t)^{1/2}(1+\delta_{\lambda^k})^{1/2}\frac{\mu_2}{1-\mu_1}\|\mathbf{W}\|_F+\big(1+\frac{1}{t}\big)^{1/2}\|\mathbf{W}\|_F\\ =&\big((1+t)^{1/2}(1+\delta_{\lambda^k})^{1/2}\frac{\mu_2}{1-\mu_1}+\big(1+\frac{1}{t}\big)^{1/2}\big)\|\mathbf{W}\|_F.\end{aligned}$$

where (a) is from the fact that $\|\mathbf{R}^{k+\lceil c\lambda^k\rceil}\|_F\leq\|\mathbf{R}^k\|_F$ and (14). Hence, we obtain the final conclusion in Theorem 1 with $C_0=(1+t)^{1/2}(1+\delta_{\lambda^k})^{1/2}\frac{\mu_2}{1-\mu_1}+\big(1+\frac{1}{t}\big)^{1/2}$

- **Case 2:** $L\geq 2$

Similar to the case when $L=1$, we conduct the energy comparison as

$$\|\mathbf{X}^{\Lambda^{k_L}}\|_F<\|\mathbf{X}^{\Lambda^k\setminus\Lambda_{L-1}^k}\|_F. \tag{18}$$

where $k_i=k+\sum_{\tau=1}^{i}\big\lceil\frac{c}{4}|\Lambda_\tau^k|\big\rceil$. Noticing that $|\Lambda_\tau^k|<2^\tau-1$, we can further estimate k_i as

$$k_L\leq k+\sum_{\tau=1}^{L}\Big\lceil\frac{c}{4}(2^\tau-1)\Big\rceil\overset{(a)}{\leq}k+\lceil c2^{L-1}\rceil-1. \tag{19}$$

where (a) is from [2, Eq. (18)].

- **Upper bound for** $\|\mathbf{X}^{\Lambda^{k_L}}\|_F$

  We have

$$\begin{aligned}\|\mathbf{R}^{k_L}\|_F=&\|\mathbf{\Phi}(\mathbf{X}-\hat{\mathbf{X}}^{k_L})+\mathbf{W}\|_F\\ &\overset{(a)}{\geq}\|\mathbf{\Phi}(\mathbf{X}-\hat{\mathbf{X}}^{k_L})\|_F-\|\mathbf{W}\|_F\\ &\geq(1-\delta_{|S\cup S^{k_L}|})^{1/2}\|\mathbf{X}-\hat{\mathbf{X}}^{k_L}\|_F-\|\mathbf{W}\|_F\\ &\overset{(b)}{\geq}(1-\delta_{|S\cup S^{k_L}|})^{1/2}\|\mathbf{X}^{\Lambda^{k_L}}\|_F-\|\mathbf{W}\|_F\\ &\overset{(c)}{\geq}(1-\delta_{p_2})^{1/2}\|\mathbf{X}^{\Lambda^{k_L}}\|_F-\|\mathbf{W}\|_F.\end{aligned}$$

  where (a) comes from the triangle inequality, (b) is based on $\mathrm{rsupp}(\mathbf{X}-\hat{\mathbf{X}}^{k_L})=S\cup S^{k_L},\mathrm{rsupp}(\hat{\mathbf{X}}^{k_L})=S^{k_L}=\Omega\setminus\Lambda^{k_L}$ and (c) is because $|S\cup S^{k_L}|\leq|\Lambda^{k_L}|+|S^{k_L}|\leq k_L+\lambda^k<k+\lceil c2^{L-1}\rceil-1+\lambda^k\leq k+\lceil(c+1)\lambda^k\rceil=p_2$ and RIC is non-decreasing.

  Through a little transformation, we obtain the upper bound as

$$\|\mathbf{X}^{\Lambda^{k_L}}\|_F\leq\frac{\|\mathbf{R}^{k_L}\|_F+\|\mathbf{W}\|_F}{(1-\delta_{p_2})^{1/2}}. \tag{20}$$

- **Lower bound for** $\|\mathbf{X}^{\Lambda^k\setminus\Lambda_{L-1}^k}\|_F$

  From multi-step residual analysis (Proposition 2), we get the following inequalities:

$$\begin{aligned}&\|\mathbf{R}^{k_1}\|_F-\|\mathbf{\Phi}_{\Lambda^k\setminus\Lambda_1^k}\mathbf{X}^{\Lambda^k\setminus\Lambda_1^k}+\mathbf{W}\|_F^2\\ &\leq C_{1,k,k_1-k}(\|\mathbf{R}^k\|_F-\|\mathbf{\Phi}_{\Lambda^k\setminus\Lambda_1^k}\mathbf{X}^{\Lambda^k\setminus\Lambda_1^k}+\mathbf{W}\|_F^2),\end{aligned} \tag{21a}$$

$$\begin{aligned}&\|\mathbf{R}^{k_2}\|_F-\|\mathbf{\Phi}_{\Lambda^k\setminus\Lambda_2^k}\mathbf{X}^{\Lambda^k\setminus\Lambda_2^k}+\mathbf{W}\|_F^2\\ &\leq C_{2,k_1,k_2-k_1}(\|\mathbf{R}^{k_1}\|_F-\|\mathbf{\Phi}_{\Lambda^k\setminus\Lambda_2^k}\mathbf{X}^{\Lambda^k\setminus\Lambda_2^k}+\mathbf{W}\|_F^2),\end{aligned} \tag{21b}$$

$$
\begin{aligned}
&\vdots\\
&\|\mathbf{R}^{k_L}\|_F - \|\mathbf{\Phi}_{\Lambda^k\backslash\Lambda_L^k}\mathbf{X}^{\Lambda^k\backslash\Lambda_L^k}+\mathbf{W}\|_F^2\\
&\le C_{L,k_{L-1},k_L-k_{L-1}}\\
&\quad\times(\|\mathbf{R}^{k_{L-1}}\|_F - \|\mathbf{\Phi}_{\Lambda^k\backslash\Lambda_L^k}\mathbf{X}^{\Lambda^k\backslash\Lambda_L^k}+\mathbf{W}\|_F^2).
\end{aligned}
\tag{21c}
$$

For any $i \in \{1, 2, \cdots, L\}$, we set a bound for the constant $C_{i,k_{i-1},k_i-k_{i-1}}$. Considering

$$
\frac{k_i-k_{i-1}}{|\Lambda_i^k|} = \frac{\lceil \frac{c}{4}|\Lambda_\tau^k|\rceil}{|\Lambda_i^k|} \ge \frac{c}{4},
$$

we can further make the estimation as

$$
\begin{aligned}
C_{i,k_{i-1},k_i-k_{i-1}} &= \exp\left(-\frac{(k_i-k_{i-1})(1-\delta_{|\Lambda_i^k\cup S^{k_i}|})^2}{\max\{1,|\Lambda_i^k\backslash S^{k_{i-1}}|\}}\right)\\
&\overset{(a)}{\le} \exp\left(-\frac{(k_i-k_{i-1})(1-\delta_{|\Lambda_i^k\cup S^{k_i}|})^2}{|\Lambda_i^k|}\right)\\
&\le \exp\big(-\frac{c}{4}(1-\delta_{|\Lambda_i^k\cup S^{k_i}|})^2\big)\\
&\overset{(b)}{\le} \exp\left(-\frac{c}{4}(1-\delta_{p_2})^2\right)\\
&= \frac{1}{2\beta}
\end{aligned}
\tag{22}
$$

where (a) is because $\max\{1,|\Lambda_i^k\backslash S^{k_{i-1}}|\} \le \max\{1,|\Lambda_i^k|\} = |\Lambda_i^k|$ and (b) is due to $|\Lambda_i^k \cup S^{k_i}| \le |\Lambda^k|+|S^{k_i}| \le \lambda^k+k_i \le \lambda^k+k_L \le \lceil(c+1)\lambda^k\rceil+k = p_2$.

Employing (21c) - (21a) in turn and define $\sigma = \frac{1}{2\beta} = \exp\left(-\frac{c}{4}(1-\delta_{p_2})^2\right)$, we can get

$$
\begin{aligned}
&\|\mathbf{R}^{k_L}\|_F^2\\
\le&\sigma^L\|\mathbf{R}^k\|_F^2+(1-\sigma)\sum_{i=1}^L\sigma^{L-i}\|\mathbf{\Phi}_{\Lambda^k\backslash\Lambda_i^k}\mathbf{X}^{\Lambda^k\backslash\Lambda_i^k}+\mathbf{W}\|_F^2\\
\overset{(a)}{\le}&\sigma^L\|\mathbf{R}^k\|_F^2+(1-\sigma)\sum_{i=1}^L\sigma^{L-i}\\
&\times\big((1+t)\|\mathbf{\Phi}_{\Lambda^k\backslash\Lambda_i^k}\mathbf{X}^{\Lambda^k\backslash\Lambda_i^k}\|_F^2+\big(1+\frac{1}{t}\big)\|\mathbf{W}\|_F^2\big)\\
\overset{(b)}{\le}&\sigma^L\|\mathbf{R}^k\|_F^2+(1-\sigma)(1+t)(1+\delta_{\lambda^k})\sum_{i=1}^L\sigma^{L-i}\\
&\times\|\mathbf{X}^{\Lambda^k\backslash\Lambda_i^k}\|_F^2+(1-\sigma)\big(1+\frac{1}{t}\big)\sum_{i=1}^L\sigma^{L-i}\|\mathbf{W}\|_F^2\\
\overset{(c)}{\le}&\sigma^L(1+t)(1+\delta_{\lambda^k})\|\mathbf{X}^{\Lambda^k}\|_F^2+(1-\sigma)(1+t)\\
&\times(1-\delta_{\lambda^k})\sum_{i=1}^L\sigma^{L-i}\|\mathbf{X}^{\Lambda^k\backslash\Lambda_i^k}\|_F^2+\big(1+\frac{1}{t}\big)\|\mathbf{W}\|_F^2\\
\overset{(d)}{\le}&\frac{1}{\beta}(1-\sigma)(1+t)(1+\delta_{\lambda^k})\sum_{i=1}^L(\beta\sigma)^{L-i}\|\mathbf{X}^{\Lambda^k\backslash\Lambda_{L-1}^k}\|_F^2\\
&+\frac{1}{\beta}(\beta\sigma)^L(1+t)(1+\delta_{\lambda^k})\|\mathbf{X}^{\Lambda^k\backslash\Lambda_{L-1}^k}\|_F^2\\
&+\big(1+\frac{1}{t}\big)\|\mathbf{W}\|_F^2\\
\overset{(e)}{=}&\frac{1}{\beta}\big(\frac{1}{2^L}+(2-\frac{2}{2^{L-1}})(1-\sigma)\big)(1+t)(1+\delta_{\lambda^k})\\
&\times\|\mathbf{X}^{\Lambda^k\backslash\Lambda_{L-1}^k}\|_F^2+\big(1+\frac{1}{t}\big)\|\mathbf{W}\|_F^2\\
\overset{(f)}{\le}&\frac{1}{\beta}\big(\frac{1}{2^{L-1}}+2-\frac{2}{2^{L-1}}\big)(1+t)(1+\delta_{\lambda^k})(1-\sigma)\\
&\times\|\mathbf{X}^{\Lambda^k\backslash\Lambda_{L-1}^k}\|_F^2+\big(1+\frac{1}{t}\big)\|\mathbf{W}\|_F^2\\
\le&4\sigma(1+t)(1+\delta_{\lambda^k})(1-\sigma)\|\mathbf{X}^{\Lambda^k\backslash\Lambda_{L-1}^k}\|_F^2\\
&+\big(1+\frac{1}{t}\big)\|\mathbf{W}\|_F^2,
\end{aligned}
$$

where (a) comes from the inequality $(a+b)^2 \le (1+t)a^2+(1+\frac{1}{t})b^2$, (b) is due to the RIP of $\mathbf{\Phi}$, (c) uses (15), (d) is base on the definition of L, (e) is from $\beta\sigma = \frac{1}{2}$ and (f) is because we can derive $\sigma < \frac{1}{2}$ from the second condition $c > c''$.
Hence, with the inequality $\sqrt{a^2+b^2} \le a+b$, we can derive that

$$
\|\mathbf{R}^{k_L}\|_F \le \tau_0\|\mathbf{X}^{\Lambda^k\backslash\Lambda_{L-1}^k}\|_F+\big(1+\frac{1}{t}\big)^{1/2}\|\mathbf{W}\|_F, \tag{23}
$$

where $\tau_0 = \beta^{-1/2}(1+t)^{1/2}(1+\delta_{\lambda^k})^{1/2}(2-\frac{1}{\beta})^{1/2}$. Hence, we get the lower bound for $\|\mathbf{X}^{\Lambda^k\backslash\Lambda_{L-1}^k}\|_F$ as

$$
\|\mathbf{X}^{\Lambda^k\backslash\Lambda_{L-1}^k}\|_F \ge \frac{1}{\tau_0}(\|\mathbf{R}^{k_L}\|_F-\big(1+\frac{1}{t}\big)^{1/2}\|\mathbf{W}\|_F).
$$

Using (20) and (23), we finally conclude that

$$
\|\mathbf{X}^{\Lambda^{k_L}}\|_F \le \tau_1\|\mathbf{X}^{\Lambda^k\backslash\Lambda_{L-1}^k}\|_F+\tau_0\|\mathbf{W}\|_F,
$$

where $\tau_1 = (1-\delta_{p_2})^{-1/2}\tau_0$ and $\tau_2 = (1-\delta_{p_2})^{-1/2}\left(1+\big(1+\frac{1}{t}\big)\right)^{1/2}$. Since the condition $c > c''$ holds true, we can easily derive $\tau_1 < 1$. Therefore, when $\|\mathbf{W}\|_F < \frac{1-\tau_1}{\tau_2}\|\mathbf{X}^{\Lambda^k\backslash\Lambda_{L-1}^k}\|_F$, (18) is obviously correct. On the other hand, when $\|\mathbf{W}\|_F \ge \frac{1-\tau_1}{\tau_2}\|\mathbf{X}^{\Lambda^k\backslash\Lambda_{L-1}^k}\|_F$, we directly prove the theorem by

$$
\begin{aligned}
&\|\mathbf{R}^{k+\lceil c\lambda^k\rceil}\|_F^2\\
\overset{(a)}{\le}&(1-\delta_{p_2})^{1/2}\tau_1\|\mathbf{X}^{\Lambda^k\backslash\Lambda_{L-1}^k}\|_F+\big(1+\frac{1}{t}\big)^{1/2}\|\mathbf{W}\|_F\\
\le&\big((1-\delta_{p_2})^{1/2}\frac{\tau_1\tau_2}{1-\tau_1}+\big(1+\frac{1}{t}\big)^{1/2}\big)\|\mathbf{W}\|_F.
\end{aligned}
$$

where (a) is because $\|\mathbf{R}^{k+\lceil c\lambda^k\rceil}\|_F^2 \le \|\mathbf{R}^{k_L}\|$ and (b) comes from (23). Hence, we prove the final conclusion with $C_0 = \big((1-\delta_{p_2})^{1/2}\frac{\tau_1\tau_2}{1-\tau_1}+\big(1+\frac{1}{t}\big)^{1/2}\big)$.

Finally, we will carry out the induction based on (12) and (18). For $L = 1$, since rows in $\mathbf{X}$ are sorted in descending order of their ℓ_2 norms, the energy comparison (12) indicates that $|\Lambda^{k+1}| < |\Lambda^k| = N$. By inductive assumption, there exists a constant C_0 such that:

$$
\|\mathbf{R}^{k+1+\lceil c\lambda^{k+1}\rceil}\|_F \le C_0\|\mathbf{W}\|_F.
$$

Further, since $c > 1$, it is obvious that

$$
\begin{aligned}
k+1+\lceil c\lambda^{k+1}\rceil &= k+\lceil 1+c\lambda^{k+1}\rceil\\
&\le k+\lceil c(\lambda^{k+1}+1)\rceil\\
&\le k+\lceil cN\rceil.
\end{aligned}
\tag{24}
$$

Hence, we acquire our conclusion by

$$\|\mathbf{R}^{k+\lceil cN \rceil}\|_F \leq \|\mathbf{R}^{k+1+\lceil c\lambda^{k+1}\rceil}\|_F \leq C_0\|\mathbf{W}\|_F.$$

For $L = 2$, we can obtain $|\Lambda^{k_L}| < |\Lambda^k \backslash \Lambda^k_{L-1}| \leq N - 2^{L-1} + 1$ from (18) and further derive

$$\begin{aligned} k_L + \lceil c\lambda^{k_L}\rceil &\overset{(19)}{\leq} k + \lceil c2^{L-1}\rceil - 1 + \lceil c(N - 2^{L-1})\rceil \\ &\leq k + \lceil cN \rceil . \end{aligned} \tag{25}$$

Similar to the analysis where $L = 1$, we have:

$$\begin{aligned} \|\mathbf{R}^{k+\lceil cN\rceil}\|_F &\leq \|\mathbf{R}^{k_L+\lceil c\lambda^{k_L}\rceil}\|_F \\ &\leq C_0\|\mathbf{W}\|_F. \end{aligned}$$

Combining these two cases, we complete the proof for Theorem 1.

E. Proof of Theorem 2

Theorem 2. *Consider the MMV model $\mathbf{Y} = \mathbf{\Phi X} + \mathbf{W}$. Then, SOMP robustly recover any joint K-sparse signal $\mathbf{X}$ with*

$$\begin{aligned} &\|\mathbf{X} - \mathbf{X}^{\lceil cK\rceil}\|_F \leq C_1\|\mathbf{W}\|_F, \\ &\|\mathbf{X} - \hat{\mathbf{X}}\|_F \leq C\,\|\mathbf{W}\|_F\,, \end{aligned}$$

if $c > \max\big\{ -\frac{4}{(1-\delta)^2}\ln\big(\frac{1}{2} - \frac{1}{2(1-\delta)^2} + \frac{1}{2(1-\delta)(1+t)(1+\delta)}\big), -\frac{4}{(1-\delta)^2}\ln\big(\frac{1}{2} - \frac{1}{2}\sqrt{1 - \frac{1-\delta}{(1+t)(1+\delta)}}\big)\big\}$ *for some constant $t > 0$, and C_1, C are constants depending on $\delta := \delta_{\lceil (c+1)K\rceil}$.*

Employing Theorem 1 with $k = 0$ and $\lambda^k = K$, we have $\|\mathbf{R}^{\lceil cK\rceil}\|_F \leq C_0\|\mathbf{W}\|_F$. On the other hand, the energy of residual can be estimated as

$$\begin{aligned} \|\mathbf{R}^{\lceil cK\rceil}\|_F &= \|\mathbf{\Phi}(\mathbf{X} - \hat{\mathbf{X}}^{\lceil cK\rceil}) + \mathbf{W}\|_F \\ &\overset{(a)}{\geq} \|\mathbf{\Phi}(\mathbf{X} - \hat{\mathbf{X}}^{\lceil cK\rceil})\|_F - \|\mathbf{W}\|_F \\ &\geq (1 - \delta_{|S\cup S^{\lceil cK\rceil}|})^{1/2}\|\mathbf{X} - \hat{\mathbf{X}}^{\lceil cK\rceil}\|_F - \|\mathbf{W}\|_F \\ &\overset{(b)}{\geq} (1 - \delta_{p_2})^{1/2}\|\mathbf{X} - \hat{\mathbf{X}}^{\lceil cK\rceil}\|_F - \|\mathbf{W}\|_F, \end{aligned}$$

where (a) comes from the triangle inequality and (b) is due to $|S \cup S^{\lceil cK\rceil}| < |S| + |S^{\lceil cK\rceil}| \leq \lceil (c+1)K\rceil = p_2$. Hence, it can be derived that

$$\begin{aligned} \|\mathbf{X} - \hat{\mathbf{X}}^{\lceil cK\rceil}\|_F \leq&(1 - \delta_{p_2})^{-1/2}(\|\mathbf{R}^{\lceil cK\rceil}\|_F + \|\mathbf{W}\|_F) \\ \leq&\frac{C_0 + 1}{(1 - \delta_{p_2})^{1/2}}\|\mathbf{W}\|_F. \end{aligned} \tag{26}$$

We complete the proof of the first conclusion with $C_1 = \frac{C_0+1}{(1-\delta_{p_2})^{1/2}}$. Now, considering the estimation after pruning, i.e., $\hat{\mathbf{X}}$, suppose that $\text{rsupp}(\hat{\mathbf{X}}) = \hat{S}$ and define $\mathbf{Z}$ as $(\mathbf{Z})^{\hat{S}} = (\hat{\mathbf{X}}^{\lceil cK\rceil})^{\hat{S}}$ and $(\mathbf{Z})^{\Omega\backslash\hat{S}} = \mathbf{0}$,then

$$\begin{aligned} &\|\hat{\mathbf{X}} - \mathbf{X}\|_F \\ &\leq \|\hat{\mathbf{X}} - \mathbf{Z}\|_F + \|\mathbf{Z} - \hat{\mathbf{X}}^{\lceil cK\rceil}\|_F + \|\hat{\mathbf{X}}^{\lceil cK\rceil} - \mathbf{X}\|_F \\ &\overset{(a)}{\leq} \|\hat{\mathbf{X}} - \mathbf{Z}\|_F + 2\|\hat{\mathbf{X}}^{\lceil cK\rceil} - \mathbf{X}\|_F \\ &\overset{(b)}{\leq} \|\hat{\mathbf{X}} - \mathbf{Z}\|_F + 2C_1\|\mathbf{W}\|_F, \end{aligned} \tag{27}$$

where (a) is from the selection criterion of $\hat{S}$ and the definition of $\mathbf{Z}$ and (b) uses (26). Next, we analyze the first term.

$$\begin{aligned} &\|\hat{\mathbf{X}} - \mathbf{Z}\|_F \\ &\leq \frac{1}{(1 - \delta_{2K})^{1/2}}\|\mathbf{\Phi}(\hat{\mathbf{X}} - \mathbf{Z})^{\hat{S}})\|_F \\ &\overset{(a)}{\leq} \frac{1}{(1 - \delta_{2K})^{1/2}}(\|\mathbf{Y} - \mathbf{\Phi}\hat{\mathbf{X}}\|_F + \|\mathbf{Y} - \mathbf{\Phi Z}\|_F) \\ &\overset{(b)}{\leq} \frac{2}{(1 - \delta_{2K})^{1/2}}\|\mathbf{\Phi}(\mathbf{X} - \mathbf{Z})\|_F \\ &\overset{(c)}{\leq} \frac{2(1 + \delta_{2K})^{1/2}}{(1 - \delta_{2K})^{1/2}}\|\mathbf{X} - \mathbf{Z}\|_F \\ &\overset{(d)}{\leq} \frac{2(1 + \delta_{2K})^{1/2}}{(1 - \delta_{2K})^{1/2}}(\|\mathbf{X} - \hat{\mathbf{X}}^{\lceil cK\rceil}\|_F + \|\hat{\mathbf{X}}^{\lceil cK\rceil} - \mathbf{Z}\|_F) \\ &\overset{(e)}{\leq} \frac{4(1 + \delta_{2K})^{1/2}}{(1 - \delta_{2K})^{1/2}}\|\mathbf{X} - \hat{\mathbf{X}}^{\lceil cK\rceil}\|_F \\ &\overset{(f)}{\leq} \frac{4C_1(1 + \delta_{2K})^{1/2}}{(1 - \delta_{2K})^{1/2}}\|\mathbf{W}\|_F, \end{aligned} \tag{28}$$

where (a) and (d) uses the triangle inequality, (b) is from the selection criterion of $\hat{\mathbf{X}}$, (c) relies on the RIP of measurement matrix $\mathbf{\Phi}$, (e) is from the definition of $\mathbf{Z}$ and $\hat{S}$ and (f) can be directly derived from (26). Finally, combining (27) and (28), we have

$$\|\hat{\mathbf{X}} - \mathbf{X}\|_F \leq \big(2C_1 + \frac{4C_1(1 + \delta_{2K})^{1/2}}{(1 - \delta_{2K})^{1/2}}\big)\|\mathbf{W}\|_F.$$

Therefore, we complete the proof of the second conclusion with $C_2 = 2C_1 + \frac{4C_1(1+\delta_{2K})^{1/2}}{(1-\delta_{2K})^{1/2}}$.

References

[1] T. Zhang, "Sparse recovery with orthogonal matching pursuit under RIP," *IEEE Trans. Inf. Theory*, vol. 57, no. 9, pp. 6215–6221, Sep. 2011.

[2] J. Wang and B. Shim, "Exact recovery of sparse signals using orthogonal matching pursuit: How many iterations do we need?," *IEEE Transactions on Signal Processing*, vol. 64, no. 16, pp. 4194–4202, 2016.