Supplementary

A. Prelininary

Before computing the proof of Proposition 1, we first introduce and prove some necessary lemmas. Lemma 1 is a straightforward deduction when we consider the projection as a type of scaling transformation along all directions.

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

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

Proof. Notice the definition that $\mathbf{P}_{\mathcal{R}(\phi)} = \phi(\phi^T\phi)^{-0.25}\phi^T$. Since $\phi^T\phi$ is a real number, we can derive that

$$\begin{aligned}\|\mathbf{P}_{\mathcal{R}(\phi)}x\|_2^2 &= \|\phi(\phi^T\phi)^{-0.25}\phi^T x\|_2^2\\ &= \frac{1}{(\phi^T\phi)^2}\|\phi\phi^T x\|_2^2\\ &= \frac{1}{(\phi^T\phi)^2}x^T\phi\phi^T\phi\phi^T x\\ &= \frac{1}{(\phi^T\phi)}x^T\phi\phi^T x\\ &= \frac{1}{\|\phi\|_2}\|\phi^T x\|_2.\end{aligned}$$

□

Next, we explore the relationship between the inner product and its derived norms.

Lemma 2. *Suppose $\mathbf{U} \in \mathbb{R}^{m\times n}$ and $\mathbf{V} \in \mathbb{R}^{m\times n}$ are two matrices with $S = rsupp(\mathbf{U}) \cup rsupp(\mathbf{V})$. Let j denote the index of the row with largest ℓ_2 norm. Then the following relationship holds true.*

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

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

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

By assumption, since $\mathbf{V}^j$ has the largest ℓ_2 norm. We can make further estimation for the Frobenious norm $\|\mathbf{V}^S\|_F$ as

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

Combining (6) with (3), we obtain the final relationships. □

Finally, an explicit representation for residual reduction of SOMP within a certain iteration is given, which is expressed in the form of the orthogonal projection.

Proposition 3. *In the k-th iteration of SOMP, the residual satisfies*

$$\|\mathbf{R}^k\|_F^2 - \|\mathbf{R}^{k+1}\|_F^2 = \|\mathbf{P}_{\mathcal{R}(\mathcal{P}^{\perp}_{S^k}\phi_{s^{k+1}})}\mathbf{R}^k\|_F^2.$$

Proof. As is shown, the residual of SOMP can be seen as the orthogonal projection of measurement vectors onto the orthogonal complement space of $\mathcal{R}(\mathbf{\Phi}_{S^k})$ (i.e, $\mathbf{R}^k = \mathbf{P}^{\perp}_{S^k}\mathbf{Y}$). Based on this property, we have

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

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

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

The residual reduction can be transformed as:

$$\begin{aligned}&\|\mathbf{R}^k\|_F^2 - \|\mathbf{R}^{k+1}\|_F^2\\ =&\langle \mathbf{R}^k - \mathbf{R}^{k+1}, \mathbf{R}^k - \mathbf{R}^{k+1}\rangle_F + 2\langle \mathbf{R}^{k+1}, \mathbf{R}^k - \mathbf{R}^{k+1}\rangle_F\\ =&\|\mathbf{R}^k - \mathbf{R}^{k+1}\|_F^2\\ =&\|\mathbf{P}_{S^{k+1}}\mathbf{R}^k\|_F^2\\ =&\|\mathbf{\Phi}_{S^{k+1}}\mathbf{\Phi}^{\dagger}_{S^{k+1}}\mathbf{R}^k\|_F^2\\ =&\|\mathbf{\Phi}_{S^{k+1}}(\mathbf{\Phi}^T_{S^{k+1}}\mathbf{\Phi}_{S^{k+1}})^{-0.25}\mathbf{\Phi}^T_{S^{k+1}}\mathbf{R}^k\|_F^2.\end{aligned} \tag{4}$$

Note that $S^{k+1} = S^k \cup s^{k+1}$, we express $\mathbf{\Phi}_{S^{k+1}}$ in the form of block matrix as $\mathbf{\Phi}_{S^{k+1}} = [\mathbf{\Phi}_{S^k}\ \ \phi_{s^{k+1}}]$ and then

$$\mathbf{\Phi}^T_{S^{k+1}}\mathbf{\Phi}_{S^{k+1}} = \begin{bmatrix}\mathbf{\Phi}^T_{S^k}\mathbf{\Phi}_{S^k} & \mathbf{\Phi}^T_{S^k}\phi_{s^{k+1}}\\ \phi^T_{s^{k+1}}\mathbf{\Phi}_{S^k} & \phi^T_{s^{k+1}}\phi_{s^{k+1}}\end{bmatrix}.$$

By the inverse formula for a block matrix, we have

$$(\mathbf{\Phi}^T_{S^{k+1}}\mathbf{\Phi}_{S^{k+1}})^{-0.25} = \begin{bmatrix}\mathbf{B} & \mathbf{C}\\ \mathbf{D} & \mathbf{E}\end{bmatrix},$$

where

$$\begin{aligned}\mathbf{B} &= (\mathbf{\Phi}^T_{S^k}(\mathbf{I} - \phi_{s^{k+1}}\phi^{\dagger}_{s^{k+1}})\mathbf{\Phi}_{S^k})^{-0.25},\\ \mathbf{E} &= (\phi^T_{s^{k+1}}(\mathbf{I} - \mathbf{\Phi}_{S^k}\mathbf{\Phi}^{\dagger}_{S^k})\phi_{s^{k+1}})^{-0.25},\\ \mathbf{C} &= -(\mathbf{\Phi}^T_{S^k}\mathbf{\Phi}_{S^k})^{-0.25}\mathbf{\Phi}^T_{S^k}\phi_{s^{k+1}}\mathbf{E},\\ \mathbf{D} &= -(\phi^T_{s^{k+1}}\phi_{s^{k+1}})^{-0.25}\phi^T_{s^{k+1}}\mathbf{\Phi}_{S^k}\mathbf{B}.\end{aligned}$$

Now we can further simplify the expression of (4) as

$$\begin{aligned}&\|\mathbf{\Phi}_{S^{k+1}}(\mathbf{\Phi}^T_{S^{k+1}}\mathbf{\Phi}_{S^{k+1}})^{-0.25}\mathbf{\Phi}^T_{S^{k+1}}\mathbf{R}^k\|_F^2.\\ \overset{(a)}{=}&\|(\mathbf{\Phi}_{S^{k+1}}(\mathbf{\Phi}^T_{S^{k+1}}\mathbf{\Phi}_{S^{k+1}})^{-0.25})_{s^{k+1}}\mathbf{\Phi}^T_{S^{k+1}}\mathbf{R}^k\|_F^2\\ =&\|(\mathbf{\Phi}_{S^k}\mathbf{C} + \phi_{s^{k+1}}\mathbf{E})\mathbf{\Phi}^T_{S^{k+1}}\mathbf{R}^k\|_F^2\\ =&\|(\mathbf{I} - \mathbf{\Phi}_{S^k}\mathbf{\Phi}^{\dagger}_{S^k})\phi_{s^{k+1}}\mathbf{E}\mathbf{\Phi}^T_{S^{k+1}}\mathbf{R}^k\|_F^2\\ =&\|\mathbf{P}^{\perp}_{\mathbf{\Phi}_{S^k}}\phi_{s^{k+1}}(\phi^T_{s^{k+1}}\mathbf{P}^{\perp}_{\mathbf{\Phi}_{S^k}}\phi_{s^{k+1}})^{-0.25}\mathbf{\Phi}^T_{S^{k+1}}\mathbf{R}^k\|_F^2\\ =&\|\mathbf{P}^{\perp}_{\mathbf{\Phi}_{S^k}}\phi_{s^{k+1}}(\phi^T_{s^{k+1}}\mathbf{P}^{\perp}_{\mathbf{\Phi}_{S^k}}\phi_{s^{k+1}})^{-0.25}(\mathbf{\Phi}^T_{S^{k+1}}\mathbf{P}^{\perp}_{\mathbf{\Phi}_{S^k}}\mathbf{R}^k)^{s^{k+1}}\|_F^2\\ =&\|\mathbf{P}^{\perp}_{\mathbf{\Phi}_{S^k}}\phi_{s^{k+1}}(\phi^T_{s^{k+1}}\mathbf{P}^{\perp}_{\mathbf{\Phi}_{S^k}}\phi_{s^{k+1}})^{-0.25}\phi^T_{s^{k+1}}\mathbf{P}^{\perp}_{\mathbf{\Phi}_{S^k}}\mathbf{R}^k\|_F^2\\ =&\|\mathbf{P}_{\mathcal{R}(\mathbf{P}^{\perp}_{\mathbf{\Phi}_{S^k}}\phi_{s^{k+1}})}\mathbf{R}^k\|_F^2.\end{aligned}$$

where (a) comes from the fact that $\mathrm{rsupp}(\mathbf{\Phi}^T_{S^{k+1}}\mathbf{R}^k) = s^{k+1}$. Therefore, We complete the proof as above. □

B. Proof of Proposition 1

We will prove Proposition 1 in two steps. Firstly, we establish a lower bound for $\|\mathbf{R}^j\|_F^2 - \|\mathbf{R}^{j+1}\|_F^2$ and then we analyze this lower bound to find the final conclusion.

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

From Proposition 3, we establish a lower bound for residual reduction.

$$
\begin{aligned}
&\|\mathbf{R}^j\|_F^2 - \|\mathbf{R}^{j+1}\|_F^2 \\
=&\|\mathbf{P}_{\mathcal{R}(\mathbf{P}_{S^j}^{\perp}\phi_{s^{j+1}})}\mathbf{R}^j\|_F^2 \\
=&\sum_{i=1}^{r}\|\mathbf{P}_{\mathcal{R}(\mathbf{P}_{S^j}^{\perp}\phi_{s^{j+1}})}r_i^j\|_2^2 \\
\overset{(a)}{=}&\frac{1}{\|\mathbf{P}_{S^j}^{\perp}\phi_{s^{j+1}}|_2}\sum_{i=1}^{r}\|(r_i^j)^T\mathbf{P}_{S^j}^{\perp}\phi_{s^{j+1}}\|_2^2 \\
=&\frac{1}{\|\mathbf{P}_{S^j}^{\perp}\phi_{s^{j+1}}\|_2}\|(\mathbf{R}^j)^T\mathbf{P}_{S^j}^{\perp}\phi_{s^{j+1}}|_2^2 \\
\overset{(b)}{\geq}&\frac{1}{\|\phi_{s^{j+1}}\|_2}\|(\mathbf{R}^j)^T\mathbf{P}_{S^j}^{\perp}\phi_{s^{j+1}}\|_2^2 \\
\overset{(c)}{=}&\max_{i\in\Omega\backslash S^j}\|(\mathbf{R}^j)^T\mathbf{P}_{S^j}^{\perp}\phi_i\|_2^2 \\
=&\max_{i\in\Omega\backslash S^j}\|(\phi_i)^T\mathbf{R}^j\|_2^2,
\end{aligned} \tag{5}
$$

where (a) comes from Lemma 1, (b) is owing to that orthogonal projection can not increase the norm of the vector and (c) is based on the selection criterion of SOMP, combined with the fact that $\mathbf{\Phi}$ is ℓ_2 normalized. After this process, we just need to analyze the lower bound instead.

2. Analysis for lower bound

Note that $\mathrm{rsupp}(\mathbf{\Phi}^T\mathbf{R}^j) = \Omega\backslash S^j$, 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 $\mathrm{rsupp}(\mathbf{H})\cap\mathrm{rsupp}(\mathbf{\Phi}^T\mathbf{R}^j) = \Lambda_\tau^k\backslash S^j$ and Lemma 2, we have

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

where (a) is because $\Lambda_\tau^k\backslash S^j\subseteq\Omega\backslash S^j$. Next, we analyze the inner product $\left\langle\mathbf{\Phi}^T\mathbf{R}^j,\mathbf{H}\right\rangle_F$ and provide a lower bound.

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

where (a) comes from $\mathrm{rsupp}(\mathbf{\Phi}^T\mathbf{R}^j) = \Omega\backslash S^j$ and $\mathrm{rsupp}(\mathbf{X}^j) = S^j$, (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 $\mathbf{\Phi}$.

Through the above analysis, we offer the following bound:

$$
\begin{aligned}
\max_{i\in\Omega\backslash S^j}\|(\phi_i)^T\mathbf{R}^j|_2 \geq& \frac{1}{\max\{1,|\Lambda_\tau^k\backslash S^j|^{1/2}\}}\left(1-\delta_{\||\Lambda_\tau^k\cup S^j|}\right) \\
&\times\left(\|\mathbf{R}^j\|_F^2-\|\mathbf{\Phi}_{\Lambda^k\backslash\Lambda_\tau^k}\mathbf{X}^{\Lambda^k\backslash\Lambda_\tau^k}+\mathbf{W}\|_F^2\right)^{1/2}.
\end{aligned} \tag{6}
$$

Combining (5) and (6), we complete the proof of Proposition 1.

C. Proof of Proposition 2

For any $j_0\in\{j,j+1,j+2,\cdots,j+\Delta j-1\}$, we employ Proposition 1 and acquire:

$$
\begin{aligned}
&\|\mathbf{R}^{j_0}\|_F^2-\|\mathbf{R}^{j_0+1}\|_F^2 \\
\geq&\frac{\left(1-\delta_{|\Lambda_\tau^k\cup S^{j_0}|}\right)^2}{\max\{1,|\Lambda_\tau^k\backslash S^{j_0}|\}}\left(\|\mathbf{R}^{j_0}\|_F^2-\|\mathbf{\Phi}_{\Lambda^k\backslash\Lambda_\tau^k}\mathbf{X}^{\Lambda^k\backslash\Lambda_\tau^k}+\mathbf{W}\|_F^2\right) \\
\overset{(a)}{\geq}&\left(1-\exp\left(-\frac{\left(1-\delta_{|\Lambda_\tau^k\cup S^{j_0}|}\right)^2}{\max\{1,|\Lambda_\tau^k\backslash S^{j_0}|\}}\right)\right) \\
&\times(\|\mathbf{R}^{j_0}\|_F^2-\|\mathbf{\Phi}_{\Lambda^k\backslash\Lambda_\tau^k}\mathbf{X}^{\Lambda^k\backslash\Lambda_\tau^k}+\mathbf{W}\|_F^2) \\
\overset{(b)}{\geq}&\left(1-\exp\left(-\frac{\left(1-\delta_{|\Lambda_\tau^k\cup S^{j+\Delta j-0.25}|}\right)^2}{\max\{1,|\Lambda_\tau^k\backslash S^j|\}}\right)\right) \\
&\times(\|\mathbf{R}^{j_0}\|_F^2-\|\mathbf{\Phi}_{\Lambda^k\backslash\Lambda_\tau^k}\mathbf{X}^{\Lambda^k\backslash\Lambda_\tau^k}+\mathbf{W}\|_F^2),
\end{aligned}
$$

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 a little transformation, we have

$$\begin{aligned}&\|\mathbf{R}^{j_0+1}\|_F^2-\|\mathbf{\Phi}_{\Lambda^k\backslash\Lambda_\tau^k}\mathbf{X}^{\Lambda^k\backslash\Lambda_\tau^k}+\mathbf{W}\|_F^2\\ \leq&\exp\left(-\frac{\left(1-\delta_{|\Lambda_\tau^k\cup S^{j+\Delta j-0.25}|}\right)^2}{\max\{1,|\Lambda_\tau^k\backslash S^j|\}}\right)\\ &\times\left(\|\mathbf{R}^{j_0}\|_F^2-\|\mathbf{\Phi}_{\Lambda^k\backslash\Lambda_\tau^k}\mathbf{X}^{\Lambda^k\backslash\Lambda_\tau^k}+\mathbf{W}\|_F^2\right).\end{aligned}$$

Finally, we combine all these inequalities from j to $j+\Delta j-1$ and then demonstrate the conclusion.

D. Proof of Theorem 1

The proof of Theorem 1 is based on mathematical induction in the number of the remaining support indices λ^k. We first consider the case when $\lambda^k=0$. Since SOMP has chosen all support indices, the conclusion is trivial. Further, assume that the conclusion holds true for $\lambda^k<N$, then it is enough to consider the case when $\lambda^k=N$.

Using the critical point as an indicator of how many indices need to be selected after k iterations, we roughly conduct energy comparison in the following two cases.

1. In the case when L = 1

In this case, $L=1$ indicates that the energy of the remaining signal $\mathbf{X}^{\Lambda^k}$ is well concentrated on the subset Λ_1^k. Therefore, we expect to show that SOMP will select the correct index in the $(k+1)$th iteration. Instead of proving this result directly, we provide a sufficient condition using energy comparison as

$$\|\mathbf{X}^{\Lambda^{k+1}}\|_F<\|\mathbf{X}^{\Lambda^k}\|_F. \tag{7}$$

To realize this goal, we employ a common technique called nest approximation. Concretely speaking, we provide an upper bound and a lower bound separately for the left and right hand side of the (7).

1.1 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

$$\|\mathbf{X}^{\Lambda^{k+1}}\|_F\leq\frac{\|\mathbf{R}^{k+1}\|_F+\|\mathbf{W}\|_F}{(1-\delta_{p_1})^{1/2}}. \tag{8}$$

1.2 Lower bound for $\|\mathbf{X}^{\Lambda^k}\|_F$

From the one-step residual analysis (Proposition 1), we have

$$\begin{aligned}&\|\mathbf{R}^k\|_F^2-\|\mathbf{R}^{k+1}\|_F^2\\ \geq&\frac{(1-\delta_{\|\Lambda_1^k\cup S^k|})^2}{\max\{1,\|\Lambda_1^k\backslash S^k|\}}(\|\mathbf{R}^k\|_F^2-\|\mathbf{\Phi}_{\Lambda^k\backslash\Lambda_1^k}\mathbf{X}^{\Lambda^k\backslash\Lambda_1^k}+\mathbf{W}\|_F^2)\\ \overset{(a)}{=}&(1-\delta_{\|\Lambda_1^k\cup S^k|})^2(\|\mathbf{R}^k\|_F^2-\|\mathbf{\Phi}_{\Lambda^k\backslash\Lambda_1^k}\mathbf{X}^{\Lambda^k\backslash\Lambda_1^k}+\mathbf{W}\|_F^2)\\ \overset{(b)}{\geq}&(1-\delta_{p_1})^2(\|\mathbf{R}^k\|_F^2-\|\mathbf{\Phi}_{\Lambda^k\backslash\Lambda_1^k}\mathbf{X}^{\Lambda^k\backslash\Lambda_1^k}+\mathbf{W}\|_F^2),\end{aligned} \tag{9}$$

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 $\|\mathbf{\Phi}_{\Lambda^k\backslash\Lambda_1^k}\mathbf{X}^{\Lambda^k\backslash\Lambda_1^k}+\mathbf{W}\|_F$. We derive that

$$\begin{aligned}\|\mathbf{R}^k\|_F^2=&\|\mathbf{P}_{S^k}^{\perp}(\mathbf{\Phi}\mathbf{X}+\mathbf{W})\|_F^2\\ \overset{(a)}{\leq}&(1+t)\|\mathbf{P}_{S^k}^{\perp}\mathbf{\Phi}\mathbf{X}\|_F^2+(1+\frac{1}{t})\|\mathbf{W}\|_F^2\\ \overset{(b)}{\leq}&(1+t)(1+\delta_{\lambda^k})\|\mathbf{X}^{\Lambda^k}\|_F^2+(1+\frac{1}{t})\|\mathbf{W}\|_F^2,\end{aligned} \tag{10}$$

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

$$\begin{aligned}&\|\mathbf{\Phi}_{\Lambda^k\backslash\Lambda_1^k}\mathbf{X}^{\Lambda^k\backslash\Lambda_1^k}+\mathbf{W}\|_F^2\\ \leq&(1+t)\|\mathbf{\Phi}_{\Lambda^k\backslash\Lambda_1^k}\mathbf{X}^{\Lambda^k\backslash\Lambda_1^k}\|_F^2+(1+\frac{1}{t})\|\mathbf{W}\|_F^2\\ \overset{(a)}{\leq}&(1+t)(1+\delta_{|\Lambda^k\backslash\Lambda_1^k|})\|\mathbf{X}^{\Lambda^k\backslash\Lambda_1^k}\|_F^2+(1+\frac{1}{t})\|\mathbf{W}\|_F^2\\ \overset{(b)}{\leq}&(1+t)(1+\delta_{\lambda^k})\frac{1}{\beta}\|\mathbf{X}^{\Lambda^k}\|_F^2+(1+\frac{1}{t})\|\mathbf{W}\|_F^2,\end{aligned} \tag{11}$$

where (a) comes from the fact that $\mathbf{\Phi}$ satisfies the RIP and rsupp$(\mathbf{X}^{\Lambda^k\backslash\Lambda_1^k})=\Lambda^k\backslash\Lambda_1^k$ and (b) depends on the definition of critical point and the monotonicity of RIC.

Combining (9), (10) and (11), we obtain

$$\begin{aligned}&\|\mathbf{R}^{k+1}\|_F^2\\ \leq&(1-(1-\delta_{p_1})^2)\|\mathbf{R}^k\|_F^2+(1-\delta_{p_1})^2\|\mathbf{\Phi}_{\Lambda^k\backslash\Lambda_1^k}\mathbf{X}^{\Lambda^k\backslash\Lambda_1^k}+\mathbf{W}\|_F^2\\ \leq&(1-(1-\delta_{p_1})^2)((1+t)(1+\delta_{\lambda^k})\|\mathbf{X}^{\Lambda^k}\|_F^2+(1+\frac{1}{t})\|\mathbf{W}\|_F^2)\\ &+(1-\delta_{p_1})^2((1+t)(1+\delta_{\lambda^k})\frac{1}{\beta}\|\mathbf{X}^{\Lambda^k}\|_F^2+(1+\frac{1}{t})\|\mathbf{W}\|_F^2)\\ =&(1+t)(1+\delta_{\lambda^k})((1-(1-\delta_{p_1})^2)+\frac{(1-\delta_{p_1})^2}{\beta})\|\mathbf{X}^{\Lambda^k}\|_F^2\\ &+(1+\frac{1}{t})\|\mathbf{W}\|_F^2.\end{aligned}$$

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

$$\|\mathbf{R}^{k+1}\|_F^2\leq\mu_0\|\mathbf{X}^{\Lambda^k}\|_F+(1+\frac{1}{t})^{1/2}\|\mathbf{W}\|_F, \tag{12}$$

where $\mu_0=(1+t)^{1/2}(1+\delta_{\lambda^k})^{1/2}\big((1-(1-\delta_{p_1})^2)+\frac{(1-\delta_{p_1})^2}{\beta}\big)^{1/2}$.

From (8) and (12), 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, (7) 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

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

where (a) is from the fact that $\|\mathbf{R}^{k+\lceil c\lambda^k \rceil}\|_F \leq \|\mathbf{R}^k\|_F$ and (9). Hence, we obtain the final conclusion in Theorem 1 with $C_0=(1+t)^{1/2}(1+\delta_{\lambda^k})^{1/2}\frac{\mu_2}{1-\mu_1}+(1+\frac{1}{t})^{1/2}$

2. In the case when $L \geq 2$

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

$$
\|\mathbf{X}^{\Lambda^{k_L}}\|_F < \|\mathbf{X}^{\Lambda^k\setminus\Lambda^k_{L-1}}\|_F. \tag{13}
$$

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

$$
k_L \leq k+\sum_{\tau=1}^{L}\left\lceil\frac{c}{4}(2^\tau-1)\right\rceil \overset{(a)}{\leq} k+\lceil c2^{L-1}\rceil-1. \tag{14}
$$

where (a) is proved by Wang and Shim.

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

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

where (a) comes from the triangle inequality, (b) is based on rsupp$(\mathbf{X}-\hat{\mathbf{X}}^{k_L})=S\cup S^{k_L}$, 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.

$$
\|\mathbf{X}^{\Lambda^{k_L}}\|_F \leq \frac{\|\mathbf{R}^{k_L}\|_F+\|\mathbf{W}\|_F}{(1-\delta_{p_2})^{1/2}}. \tag{15}
$$

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

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

$$
\begin{aligned}
&\|\mathbf{R}^{k_1}\|_F-\|\mathbf{\Phi}_{\Lambda^k\setminus\Lambda^k_1}\mathbf{X}^{\Lambda^k\setminus\Lambda^k_1}+\mathbf{W}\|_F^2\\
&\leq C_{1,k,k_1-k}(\|\mathbf{R}^k\|_F-\|\mathbf{\Phi}_{\Lambda^k\setminus\Lambda^k_1}\mathbf{X}^{\Lambda^k\setminus\Lambda^k_1}+\mathbf{W}\|_F^2), \quad (16a)\\
&\|\mathbf{R}^{k_2}\|_F-\|\mathbf{\Phi}_{\Lambda^k\setminus\Lambda^k_2}\mathbf{X}^{\Lambda^k\setminus\Lambda^k_2}+\mathbf{W}\|_F^2\\
&\leq C_{2,k_1,k_2-k_1}(\|\mathbf{R}^{k_1}\|_F-\|\mathbf{\Phi}_{\Lambda^k\setminus\Lambda^k_2}\mathbf{X}^{\Lambda^k\setminus\Lambda^k_2}+\mathbf{W}\|_F^2), \quad (16b)\\
&\vdots\\
&\|\mathbf{R}^{k_L}\|_F-\|\mathbf{\Phi}_{\Lambda^k\setminus\Lambda^k_L}\mathbf{X}^{\Lambda^k\setminus\Lambda^k_L}+\mathbf{W}\|_F^2\\
&\leq C_{L,k_{L-1},k_L-k_{L-1}}(\|\mathbf{R}^{k_{L-1}}\|_F-\|\mathbf{\Phi}_{\Lambda^k\setminus\Lambda^k_L}\mathbf{X}^{\Lambda^k\setminus\Lambda^k_L}+\mathbf{W}\|_F^2), \quad (16c)
\end{aligned}
$$

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

$$
\frac{k_i-k_{i-1}}{|\Lambda^k_i|}=\frac{\lceil\frac{c}{4}|\Lambda^k_\tau|\rceil}{|\Lambda^k_i|}\geq\frac{c}{4},
$$

we can further make the estimation as

$$
\begin{aligned}
C_{i,k_{i-1},k_i-k_{i-1}} &= \exp\left(-\frac{\left(k_i-k_{i-1}\left(1-\delta_{|\Lambda^k_i\cup S^{k_i}|}\right)\right)^2}{\max\left\{1,|\Lambda^k_i\setminus S^{k_{i-1}}|\right\}}\right)\\
&\overset{(a)}{\leq}\exp\left(-\frac{\left(k_i-k_{i-1}\left(1-\delta_{|\Lambda^k_i\cup S^{k_i}|}\right)\right)^2}{|\Lambda^k_i\setminus S^{k_{i-1}}|}\right)\\
&\leq\exp\big(-\frac{c}{4}(1-\delta_{|\Lambda^k_i\cup S^{k_i}|})^2\big)\\
&\overset{(b)}{\leq}\exp\left(-\frac{c}{4}(1-\delta_{p_2})^2\right)\\
&=\frac{1}{2\beta}
\end{aligned} \tag{17}
$$

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

Employing (16c) - (16a) in turn and define $\sigma=\frac{1}{2\beta}=\exp\left(-\frac{c}{4}(1-\delta_{p_2})^2\right)$, we can get

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

$$\times \|\mathbf{X}^{\Lambda^k\setminus\Lambda^k_{L-1}}\|_F^2 + (1+\frac{1}{t})\|\mathbf{W}\|_F^2$$
$$\overset{(f)}{\leq}\frac{1}{\beta}(\frac{1}{2^{L-1}}+2-\frac{2}{2^{L-1}})(1+t)(1+\delta_{\lambda^k})(1-\sigma)$$
$$\times \|\mathbf{X}^{\Lambda^k\setminus\Lambda^k_{L-1}}\|_F^2 + (1+\frac{1}{t})\|\mathbf{W}\|_F^2$$
$$\leq 4\sigma(1+t)(1+\delta_{\lambda^k})(1-\sigma)\|\mathbf{X}^{\Lambda^k\setminus\Lambda^k_{L-1}}\|_F^2$$
$$+ (1+\frac{1}{t})\|\mathbf{W}\|_F^2,$$

where (a) comes from the inequality $(a+b)^2 \leq (1+t)a^2 + (1+\frac{1}{t})b^2$, (b) is due to the RIP of $\mathbf{\Phi}$, (c) uses (10), (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 $\sqrt{a^2+b^2} \leq a+b$, we can give a lower bound for $\|\mathbf{X}^{\Lambda^k\setminus\Lambda^k_{L-1}}\|_F$ as

$$\|\mathbf{R}^{k_L}\|_F \leq \tau_0\|\mathbf{X}^{\Lambda^k\setminus\Lambda^k_{L-1}}\|_F + (1+\frac{1}{t})^{1/2}\|\mathbf{W}\|_F, \quad (18)$$

where $\tau_0 = \beta^{-1/2}(1+t)^{1/2}(1+\delta_{\lambda^k})^{1/2}(2-\ \frac{1}{\beta})^{1/2}$.

Using the upper bound (15) and the lower bound (18), we finally conclude that

$$\|\mathbf{X}^{\Lambda^{k_L}}\|_F \leq \tau_1\|\mathbf{X}^{\Lambda^k\setminus\Lambda^k_{L-1}}\|_F + \tau_0\|\mathbf{W}\|_F,$$

where $\tau_1 = (1-\delta_{p_2})^{-1/2}\tau_0$ and $\tau_2 = (1-\delta_{p_2})^{-1/2}\left(1+\left(1+\frac{1}{t}\right)\right)^{1/2}$. Since the condition $c > c''$ holds true, we can easily derive $\tau_1 < 1$. Therefore, when $\|\mathbf{W}\|_F < \frac{1-\tau_1}{\tau_2}\|\mathbf{X}^{\Lambda^k\setminus\Lambda^k_{L-1}}\|_F$, (13) is obviously correct. On the other hand, when $\|\mathbf{W}\|_F \geq \frac{1-\tau_1}{\tau_2}\|\mathbf{X}^{\Lambda^k\setminus\Lambda^k_{L-1}}\|_F$, we directly prove the theorem by

$$\|\mathbf{R}^{k+\lceil c\lambda^k\rceil}\|_F^2$$
$$\overset{(a)}{\leq}(1-\delta_{p_2})^{1/2}\tau_1\|\mathbf{X}^{\Lambda^k\setminus\Lambda^k_{L-1}}\|_F + (1+\frac{1}{t})^{1/2}\|\mathbf{W}\|_F$$
$$\leq\Big((1-\delta_{p_2})^{\frac{1}{2}}\frac{\tau_1\tau_2}{1-\tau_1} + (1+\frac{1}{t})^{1/2}\Big)\|\mathbf{W}\|_F.$$

where (a) is because $\|\mathbf{R}^{k+\lceil c\lambda^k\rceil}\|_F^2 \leq \|\mathbf{R}^{k_L}\|$ and (b) comes from (18). Hence, we prove the final conclusion with $C_0 = \big((1-\delta_{p_2})^{\frac{1}{2}}\frac{\tau_1\tau_2}{1-\tau_1} + (1+\frac{1}{t})^{1/2}\big)$.

Finally, we will carry out the induction based on (7) and (13). For $L=1$, since rows in $\mathbf{X}$ are sorted in descending order of their ℓ_2 norms, the energy comparison (7) indicates that $|\Lambda^{k+1}| < |\Lambda^k| = N$. By inductive assumption, there exists a constant C_0 such that:

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

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

$$\begin{aligned} k+1+\lceil c\lambda^{k+1}\rceil &= k + \lceil 1 + c\lambda^{k+1}\rceil \\ &\leq k + \lceil c(\lambda^{k+1}+1)\rceil \\ &\leq k + \lceil cN\rceil\,. \end{aligned} \quad (19)$$

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 $|\Lambda^{k_L}| < |\Lambda^k\setminus\Lambda^k_{L-1}| \leq N - 2^{L-1}+1$ from (13) and further derive

$$\begin{aligned} k_L + \lceil c\lambda^{k_L}\rceil &\overset{(14)}{\leq} k + \lceil c2^{L-1}\rceil - 1 + \lceil c(N-2^{L-1})\rceil \\ &\leq k + \lceil cN\rceil\,. \end{aligned} \quad (20)$$

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

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

$$\begin{aligned} \|\mathbf{X}-\hat{\mathbf{X}}^{\lceil cK\rceil}\|_F &\leq (1-\delta_{p_2})^{-1/2}(\|\mathbf{R}^{\lceil cK\rceil}\|_F + \|\mathbf{W}\|_F) \\ &\leq \frac{C_0+1}{(1-\delta_{p_2})^{1/2}}\|\mathbf{W}\|_F. \end{aligned} \quad (21)$$

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 $\mathrm{rsupp}(\hat{\mathbf{X}}) = \hat{S}$ and define $\mathbf{Z}$ as $(\mathbf{Z})^{\hat{S}} = (\hat{\mathbf{X}}^{\lceil cK\rceil})^{\hat{S}}$ and $(\mathbf{Z})^{\Omega\setminus\hat{S}} = \mathbf{0}$,then

$$\begin{aligned} &\|\hat{\mathbf{X}}-\mathbf{X}\|_F \\ &\leq \|\hat{\mathbf{X}}-\mathbf{Z}\|_F + \|\mathbf{Z}-\hat{\mathbf{X}}^{\lceil cK\rceil}\|_F + \|\hat{\mathbf{X}}^{\lceil cK\rceil}-\mathbf{X}\|_F \\ &\overset{(a)}{\leq} \|\hat{\mathbf{X}}-\mathbf{Z}\|_F + 2\|\hat{\mathbf{X}}^{\lceil cK\rceil}-\mathbf{X}\|_F \\ &\overset{(b)}{\leq} \|\hat{\mathbf{X}}-\mathbf{Z}\|_F + 2C_1\|\mathbf{W}\|_F, \end{aligned} \quad (22)$$

where (a) is from the selection criterion of $\hat{S}$ and the definition of $\mathbf{Z}$ and (b) uses (21). Next, we analyze the first term.

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

$$
\begin{aligned}
&\overset{(e)}{\leq} \frac{4(1+\delta_{2K})^{1/2}}{(1-\delta_{2K})^{1/2}}\|\mathbf{X}-\hat{\mathbf{X}}^{\lceil cK \rceil}\|_F \\
&\overset{(f)}{\leq} \frac{4C_1(1+\delta_{2K})^{1/2}}{(1-\delta_{2K})^{1/2}}\|\mathbf{W}\|_F,
\end{aligned} \tag{23}
$$

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 (21). Finally, combining (22) and (23), 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 $C_2 = 2C_1 + \frac{4C_1(1+\delta_{2K})^{1/2}}{(1-\delta_{2K})^{1/2}}$.