Supplementary

A. Propositions for proof

Before computing the proof of Lemma 1, we first introduce and prove some necessary propositions. Proposition 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 satisfies*

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

Proof. Notice the definition that $\mathcal{P}_{\mathcal{R}(\phi)} = \phi(\phi^T\phi)^{-1}\phi^T$. Since $\phi^T\phi$ is a real number, we can derive that

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

□

Next, we will introduce the relationship between inner product and 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 denotes the index of the row which has the largest ℓ_2 norm. Then it satisfies*

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

Proof. From the definition of S, we can easily get $\mathbf{U}^i\mathbf{V}^i = 0$, if $i \notin S$, where $\mathbf{U}^i, \mathbf{V}^i$ denote the ith row of $\mathbf{U}, \mathbf{V}$. Therefore, through Cauthy-Schwarz inequality, we have

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

By assumption, j denotes the index of the row which has the largest ℓ_2 norm. Combined this assumption with (2), we can derive

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

Using (2) and (3), the final conclusion can be easily proved.

□

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

Proposition 1. *In the k-th iteration of SOMP, the residual satisfies*

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

Proof. As is shown, the residual of SOMP can be seen as the 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) is because $\mathbf{\Phi}\mathbf{X}^k = \mathbf{\Phi}_{S^k}\mathbf{X}^{S^k} \in \mathcal{R}(\mathbf{\Phi}_{S^k})$ and $S^K \subseteq S^{k+1}$.

Therefore, we conclude that $\left\langle \mathbf{R}^{k+1}, \mathbf{R}^k - \mathbf{R}^{k+1}\right\rangle_F = 0$ holds true since $\mathbf{R}^{k+1} = \mathbf{P}^{\perp}_{S^{k+1}}\mathbf{R}^k$ and

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

Hence, we derive

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

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

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

Employing the inverse formula for a block matrix, the inverse of $\mathbf{\Phi}_{S^{k+1}}$ is

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

where

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

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

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

where (a) is from the fact that $\text{rsupp}(\mathbf{\Phi}_{S^{k+1}}^T\mathbf{R}^k)=s^{k+1}$. We complete the proof as above. □

B. Proof of Lemma 1

With the above propositions, we begin to prove Lemma 1 in two steps.

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

From Proposition 2, we establish a lower bound for residual reduction.

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

where (a) is from Proposition 1, (b) is because orthogonal projection will not increase the norm of the vector and (c) is due to the selection criterion of SOMP and the fact that $\mathbf{\Phi}$ is ℓ_2 normalized. In this case, we just need to analyze this lower bound instead.

2. Analysis for lower bound

Note that $\text{rsupp}(\mathbf{\Phi}^T\mathbf{R}^j)=\Omega\backslash S^j$, we define the matrix $\mathbf{H}\in\mathbb{R}^{n\times r}$ 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, it is obvious that $\text{rsupp}(\mathbf{H})\cap\text{rsupp}(\mathbf{\Phi}^T\mathbf{R}^j)=\Lambda_\tau^k\backslash S^j$. By Proposition 2, we have

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

where (a) is because $\Lambda_\tau^k\backslash S^j\subseteq\Omega\backslash S^j$. Next, we analyze the inner product and provide a lower bound.

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

where (a) comes from $\text{rsupp}(\mathbf{\Phi}^T\mathbf{R}^j)=\Omega\backslash S^j,\text{rsupp}(\mathbf{X}^j)=S^j$, (b) is the property of adjoint operator, (c) is based on the cosine law, (d) is from the fact that $\mathbf{R}^j=\mathbf{\Phi}(\mathbf{X}-\hat{\mathbf{X}}^j)$, (e) is the fundamental inequality and (f) uses the RIP of the measurement matrix $\mathbf{\Phi}$.

Through this analysis, the lower bound can be bounded as

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

The final conclusion will be easily acquired by combining the above two steps together. Therefore, we complete the proof of Lemma 1.

C. Proof of Corollary 1

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

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

where (a) is 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

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

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

D. Proof of Theorem 1

The proof of Theorem 1 is based on mathematical induction in the number of 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 when $\lambda^k < N$, it is enough to consider the case when $\lambda^k = N$.

Using the inflection point of energy as an indicator of how many indices are necessary after k iterations, we can 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 $\mathbf{R}^k$ is concentrated on one dimension. 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 prove a sufficient condition using energy comparison as

$$\left\|\mathbf{X}^{\Lambda^{k+1}}\right\|_F < \left\|\mathbf{X}^{\Lambda^k}\right\|_F. \tag{5}$$

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 inequality.

1.1 Upper bound for $\left\|\mathbf{X}^{\Lambda^{k+1}}\right\|_F$

Since $\mathbf{R}^{k+1} = \mathbf{\Phi}(\mathbf{X}-\hat{\mathbf{X}}^{k+1})$,

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

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

$$\left\|\mathbf{X}^{\Lambda^{k+1}}\right\|_F \leq \frac{\left\|\mathbf{R}^{k+1}\right\|_F+\|\mathbf{W}\|_F}{(1-\delta_{p_1})^{\frac{1}{2}}}. \tag{6}$$

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

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

$$\begin{aligned}
&\left\|\mathbf{R}^k\right\|_F^2-\left\|\mathbf{R}^{k+1}\right\|_F^2\\
\geq&\frac{\left(1-\delta_{\left|\Lambda_1^k\cup S^k\right|}\right)^2}{\max\left\{1,\left|\Lambda_1^k\backslash S^k\right|\right\}}\left(\left\|\mathbf{R}^k\right\|_F^2-\left\|\mathbf{\Phi}_{\Lambda^k\backslash\Lambda_1^k}\mathbf{X}^{\Lambda^k\backslash\Lambda_1^k}+\mathbf{W}\right\|_F^2\right)\\
\overset{(a)}{=}&\left(1-\delta_{\left|\Lambda_1^k\cup S^k\right|}\right)^2\left(\left\|\mathbf{R}^k\right\|_F^2-\left\|\mathbf{\Phi}_{\Lambda^k\backslash\Lambda_1^k}\mathbf{X}^{\Lambda^k\backslash\Lambda_1^k}+\mathbf{W}\right\|_F^2\right)\\
\overset{(b)}{\geq}&(1-\delta_{p_1})^2\left(\left\|\mathbf{R}^k\right\|_F^2-\left\|\mathbf{\Phi}_{\Lambda^k\backslash\Lambda_1^k}\mathbf{X}^{\Lambda^k\backslash\Lambda_1^k}+\mathbf{W}\right\|_F^2\right).
\end{aligned}\tag{7}$$

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

It is enough to conduction 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

$$\begin{aligned}
\left\|\mathbf{R}^k\right\|_F^2&=\left\|\mathbf{P}_{S^k}^{\perp}(\mathbf{\Phi}\mathbf{X}+\mathbf{W})\right\|_F^2\\
&\overset{(a)}{\leq}(1+t)\left\|\mathbf{P}_{S^k}^{\perp}\mathbf{\Phi}\mathbf{X}\right\|_F^2+\left(1+\frac{1}{t}\right)\|\mathbf{W}\|_F^2\\
&\overset{(b)}{\leq}(1+t)(1+\delta_{\lambda^k})\left\|\mathbf{X}^{\Lambda^k}\right\|_F^2+\left(1+\frac{1}{t}\right)\|\mathbf{W}\|_F^2,
\end{aligned}\tag{8}$$

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$(\mathbf{P}_{S^k}^{\perp}\mathbf{\Phi}\mathbf{X}) = \Lambda^k$.

$$\begin{aligned}
&\left\|\mathbf{\Phi}_{\Lambda^k\backslash\Lambda_1^k}\mathbf{X}^{\Lambda^k\backslash\Lambda_1^k}+\mathbf{W}\right\|_F^2\\
\leq&(1+t)\left\|\mathbf{\Phi}_{\Lambda^k\backslash\Lambda_1^k}\mathbf{X}^{\Lambda^k\backslash\Lambda_1^k}\right\|_F^2+\left(1+\frac{1}{t}\right)\|\mathbf{W}\|_F^2\\
\overset{(a)}{\leq}&(1+t)(1+\delta_{\left|\Lambda^k\backslash\Lambda_1^k\right|})\left\|\mathbf{X}^{\Lambda^k\backslash\Lambda_1^k}\right\|_F^2+\left(1+\frac{1}{t}\right)\|\mathbf{W}\|_F^2\\
\overset{(b)}{\leq}&(1+t)(1+\delta_{\lambda^k})\frac{1}{\beta}\left\|\mathbf{X}^{\Lambda^k}\right\|_F^2+\left(1+\frac{1}{t}\right)\|\mathbf{W}\|_F^2,
\end{aligned}\tag{9}$$

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

Combining (7),(8) and (9), we obtain

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

Recall that $\beta = \frac{1}{2}\exp\left(\frac{c}{4}(1-\delta_{p_2})^2\right)$, using the conclusion $\sqrt{a^2+b^2} \leq a+b$ yield:

$$\begin{aligned}
\|\mathbf{R}^{k+1}\|_F^2 \leq&(1+t)^{\frac{1}{2}}(1+\delta_{\lambda^k})^{\frac{1}{2}}\left((1-(1-\delta_{p_1})^2)+\frac{(1-\delta_{p_1})^2}{\beta}\right)^{\frac{1}{2}}\\
&\times\|\mathbf{X}^{\Lambda^k}\|_F+(1+\frac{1}{t})^{\frac{1}{2}}\|\mathbf{W}\|_F. \qquad (10)
\end{aligned}$$

From (6) and (10), 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 $\mu_1 = (1-\delta_{p_1})^{-\frac{1}{2}}(1+t)^{\frac{1}{2}}(1+\delta_{\lambda^k})^{\frac{1}{2}}(1-(1-2\sigma)(1-\delta_{p_1})^2)^{\frac{1}{2}}$ and $\mu_2 = (1-\delta_{p_1})^{-\frac{1}{2}}\left(1+(1+\frac{1}{t})^{\frac{1}{2}}\right)$, if we define $\sigma = \exp\left(-\frac{c}{4}(1-\delta_{p_2})^2\right)$.

The first condition $c > c'$ ensures that $\mu_1 < 1$. Therefore, (5) 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

$$\begin{aligned}
&\|\mathbf{R}^{k+\max\{\lambda^k,\lfloor c\lambda^k\rfloor\}}\|_F^2\\
\overset{(a)}{\leq}&(1+t)^{\frac{1}{2}}(1+\delta_{\lambda^k})^{\frac{1}{2}}\|\mathbf{X}^{\Lambda^k}\|_F+(1+\frac{1}{t})^{\frac{1}{2}}\|\mathbf{W}\|_F\\
\leq&(1+t)^{\frac{1}{2}}(1+\delta_{\lambda^k})^{\frac{1}{2}}\frac{\mu_2}{1-\mu_1}\|\mathbf{W}\|_F+(1+\frac{1}{t})^{\frac{1}{2}}\|\mathbf{W}\|_F\\
=&\left((1+t)^{\frac{1}{2}}(1+\delta_{\lambda^k})^{\frac{1}{2}}\frac{\mu_2}{1-\mu_1}+(1+\frac{1}{t})^{\frac{1}{2}}\right)\|\mathbf{W}\|_F.
\end{aligned}$$

where (a) is from the fact that $\|\mathbf{R}^{k+\max\{\lambda^k,\lfloor c\lambda^k\rfloor\}}\|_F \leq \|R^k\|_F$ and (7).

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

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

$$\left\|\mathbf{X}^{\Lambda^{k_L}}\right\|_F < \left\|\mathbf{X}^{\Lambda^k\backslash\Lambda_{L-1}^k}\right\|_F. \qquad (11)$$

where $k_i = k + \frac{c}{4}\sum_{\tau=1}^{i}\left\lceil\left|\Lambda_\tau^k\right|\right\rceil$ and we can easily derive from the definition of Λ_τ^k that $k_i \leq k+\frac{c}{4}\sum_{\tau=1}^{i}2^{\tau-1} < k+c2^{i-2}$.

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

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

where (a) comes from the triangle inequality, (b) is based on $\text{rsupp}(\mathbf{X}-\hat{\mathbf{X}}^{k_L}) = S\cup S^{k_L}, \text{rsupp}(\mathbf{X}^{k_L}) = S^{k_L} = \Omega\backslash\Lambda^{k_L}$ and (c) is because $\left|S\cup S^{k_L}\right| \leq \left|\Lambda^{k_L}\right| + \left|S^{k_L}\right| \leq k_L+\lambda^k < k+2^{L-2}c+\lambda^k \leq k+(c+1)\lambda^k = p_2$ and RIC is non-decreasing.

Through a little transformation, we obtain the upper bound.

$$\left\|\mathbf{X}^{\Lambda^{k_L}}\right\|_F \leq \frac{\left\|\mathbf{R}^{k_L}\right\|_F+\|\mathbf{W}\|_F}{(1-\delta_{p_2})^{\frac{1}{2}}}. \qquad (12)$$

2.2 Lower bound for $\left\|\mathbf{X}^{\Lambda^k\backslash\Lambda_{L-1}^k}\right\|_F$

From multi-step residual analysis (Corollary 1), we have

$$\begin{aligned}
&\left\|\mathbf{R}^{k_1}\right\|_F-\left\|\mathbf{\Phi}_{\Lambda^k\backslash\Lambda_1^k}\mathbf{X}^{\Lambda^k\backslash\Lambda_1^k}+\mathbf{W}\right\|_F^2\\
&\leq C_{1,k,k_1-k}\left(\left\|\mathbf{R}^k\right\|_F-\left\|\mathbf{\Phi}_{\Lambda^k\backslash\Lambda_1^k}\mathbf{X}^{\Lambda^k\backslash\Lambda_1^k}+\mathbf{W}\right\|_F^2\right), \qquad (13.a)
\end{aligned}$$

$$\begin{aligned}
&\left\|\mathbf{R}^{k_2}\right\|_F-\left\|\mathbf{\Phi}_{\Lambda^k\backslash\Lambda_2^k}\mathbf{X}^{\Lambda^k\backslash\Lambda_2^k}+\mathbf{W}\right\|_F^2\\
&\leq C_{2,k_1,k_2-k_1}\left(\left\|\mathbf{R}^{k_1}\right\|_F-\left\|\mathbf{\Phi}_{\Lambda^k\backslash\Lambda_2^k}\mathbf{X}^{\Lambda^k\backslash\Lambda_2^k}+\mathbf{W}\right\|_F^2\right), \qquad (13.b)
\end{aligned}$$

$$\vdots$$

$$\begin{aligned}
&\left\|\mathbf{R}^{k_L}\right\|_F-\left\|\mathbf{\Phi}_{\Lambda^k\backslash\Lambda_L^k}\mathbf{X}^{\Lambda^k\backslash\Lambda_L^k}+\mathbf{W}\right\|_F^2\\
&\leq C_{L,k_{L-1},k_L-k_{L-1}}\left(\left\|\mathbf{R}^{k_{L-1}}\right\|_F-\left\|\mathbf{\Phi}_{\Lambda^k\backslash\Lambda_L^k}\mathbf{X}^{\Lambda^k\backslash\Lambda_L^k}+\mathbf{W}\right\|_F^2\right), \qquad (13.c)
\end{aligned}$$

For any $i \in \{1,2,\cdots,L\}$, the constant $C_{i,k_{i-1},k_i-k_{i-1}}$ is bounded by

$$\begin{aligned}
C_{i,k_{i-1},k_i-k_{i-1}} =& \exp\left(-\frac{c\left\lceil\left|\Lambda_i^k\right|\right\rceil\left(1-\delta_{\left|\Lambda_i^k\cup S^{k_{i-1}}\right|}\right)^2}{4\max\left\{1,\left|\Lambda_i^k\backslash S^{k_{i-1}}\right|\right\}}\right)\\
\overset{(a)}{\leq}&\exp\left(-\frac{c\left(1-\delta_{\left|\Lambda_i^k\cup S^{k_{i-1}}\right|}\right)^2}{4}\right)\\
\overset{(b)}{\leq}&\exp\left(-\frac{c}{4}(1-\delta_{p_2})^2\right)\\
=&\sigma \qquad (14)
\end{aligned}$$

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

Employing (13.c) - (13.a) in turn, we can get

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

where (a) is 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) comes from (8), (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 give a lower bound for $\|\mathbf{X}^{\Lambda^k\backslash\Lambda_{L-1}^k}\|_F$:

$$\begin{aligned}
\|\mathbf{R}^{k_L}\|_F \le& 2\sigma^{\frac{1}{2}}(1+t)^{\frac{1}{2}}(1+\delta_{\lambda^k})^{\frac{1}{2}}(1-\sigma)^{\frac{1}{2}}\|\mathbf{X}^{\Lambda^k\backslash\Lambda_{L-1}^k}\|_F\\
&+(1+\frac{1}{t})^{\frac{1}{2}}\|\mathbf{W}\|_F.
\end{aligned}\tag{15}$$

Using the upper bound (12) and the lower bound (15), we finally conclude that

$$\|\mathbf{X}^{\Lambda^{k_L}}\|_F \le \mu_3\|\mathbf{X}^{\Lambda^k\backslash\Lambda_{L-1}^k}\|_F + \mu_4\|\mathbf{W}\|_F,$$

where $\mu_3 = 2(1-\delta_{p_2})^{-\frac{1}{2}}\sigma^{\frac{1}{2}}(1+t)^{\frac{1}{2}}(1+\delta_{\lambda^k})^{\frac{1}{2}}(1-\sigma)^{\frac{1}{2}}$ and $\mu_4 = (1-\delta_{p_2})^{-\frac{1}{2}}\left(1+\left(1+\frac{1}{t}\right)\right)^{\frac{1}{2}}$. Since the condition $c > c''$ holds true, we can easily derive $\mu_3 < 1$.

When $\|\mathbf{W}\|_F < \frac{1-\mu_3}{\mu_4}\|\mathbf{X}^{\Lambda^k\backslash\Lambda_{L-1}^k}\|_F$, (11) is obviously correct. On the other hand, when $\|\mathbf{W}\|_F \ge \frac{1-\mu_3}{\mu_4}\|\mathbf{X}^{\Lambda^k\backslash\Lambda_{L-1}^k}\|_F$, we directly prove the theorem by

$$\begin{aligned}
&\|\mathbf{R}^{k+\max\left\{\lambda^k,\left\lfloor c\lambda^k\right\rfloor\right\}}\|_F^2\\
\overset{(a)}{\le}&(1-\delta_{p_2})^{\frac{1}{2}}\mu_3\|\mathbf{X}^{\Lambda^k\backslash\Lambda_{L-1}^k}\|_F + \left(1+\frac{1}{t}\right)^{\frac{1}{2}}\|\mathbf{W}\|_F\\
\le&((1-\delta_{p_2})^{\frac{1}{2}}\frac{\mu_3\mu_4}{1-\mu_3} + \left(1+\frac{1}{t}\right)^{\frac{1}{2}})\|\mathbf{W}\|_F.
\end{aligned}$$

where (a) is because $\|\mathbf{R}^{k+\max\left\{\lambda^k,\left\lfloor c\lambda^k\right\rfloor\right\}}\|_F^2 \le \|\mathbf{R}^{k_L}$ and (b) comes from (15).

Finally, we will carry out the induction based on (5) and (11). For $L = 1$, since rows in $\mathbf{X}$ are sorted in descending order of their ℓ_2 norms, the energy comparison (5) indicates that $|\Lambda^{k+1}| < |\Lambda^k| = N$. By inductive assumption, there exists a constant c_{k+1} such that:

$$\|\mathbf{R}^{k+1+\max\left\{\lambda^{k+1},\left\lfloor c\lambda^{k+1}\right\rfloor\right\}}\|_F \le c_{k+1}\|\mathbf{W}\|_F.$$

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

$$\begin{aligned}
&k+1+\max\left\{\lambda^{k+1},\left\lfloor c\lambda^{k+1}\right\rfloor\right\}\\
=&k+\max\left\{1+\lambda^{k+1},1+\left\lfloor c\lambda^{k+1}\right\rfloor\right\}\\
\le&k+\max\left\{1+\lambda^{k+1},\left\lfloor c(\lambda^{k+1}+1)\right\rfloor\right\}\\
\le&k+\max\left\{\lambda^k,\left\lfloor c\lambda^k\right\rfloor\right\},
\end{aligned}$$

Hence, we acquire our conclusion by

$$\begin{aligned}
&\left\|\mathbf{R}^{k+\max\left\{\lambda^k,\left\lfloor c\lambda^k\right\rfloor\right\}}\right\|_F\\
\le&\left\|\mathbf{R}^{k+1+\max\left\{\lambda^{k+1},\left\lfloor c\lambda^{k+1}\right\rfloor\right\}}\right\|_F\\
\le&c_k\left\|\mathbf{W}\right\|_F,
\end{aligned}$$

where $c_k = 1 + c_{k+1}$ is a constant determined by RIC only.

For $L = 2$, we can obtain $\left|\Lambda^{k_L}\right| < \left|\Lambda^k\backslash\Lambda_{L-1}^k\right| \le \lambda^k - 2^{L-2}$ from (11) and yield

$$\begin{aligned}
&k_L + \max\left\{\lambda^{k_L},\left\lfloor c\lambda^{k_L}\right\rfloor\right\}\\
\le& k + 2^{L-2}c + \max\left\{\lambda^k - 2^{L-2},\left\lfloor c(\lambda^k - 2^{L-2} + 2^{L-2})\right\rfloor\right\}\\
=& k + \max\left\{\lambda^k - 2^{L-2} + 2^{L-2}c,\left\lfloor c\lambda^k\right\rfloor\right\}\\
\le& k + \max\left\{\lambda^k,\left\lfloor c\lambda^k\right\rfloor\right\},
\end{aligned}$$

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

$$\begin{aligned}
&\left\|\mathbf{R}^{k+\max\left\{\lambda^k,\left\lfloor c\lambda^k\right\rfloor\right\}}\right\|_F\\
\le&\left\|\mathbf{R}^{k_L+\max\left\{\lambda^{k_L},\left\lfloor c\lambda^{k_L}\right\rfloor\right\}}\right\|_F\\
\le&c_k\left\|\mathbf{W}\right\|_F,
\end{aligned}$$

where $c_k = 1 + c_{k_L}$ is a constant determined by RIC only.