Supplementary: Boundary Multiple Measurement Vectors for Multi-Coset Sampler

Dong Xiao, Jian Wang and Yun Lin

This supplementary material is dedicated to the proofs for Theorems 1–3 in our main paper.

Before proceeding to the proofs, we review some useful notations. For a complex matrix $\mathbf{X} \in \mathbb{C}^{n\times L}$ and a set $S \subseteq \{1,\cdots,n\}$, $\mathbf{X}_S$ (or $\mathbf{X}^S$) denotes the submatrix of $\mathbf{X}$ with columns (or rows) indexed by S; $\mathbf{X}_{i,j}$, $\mathbf{X}_{i,:}$ and $\mathbf{X}_{:,i}$ are the (i,j)th entry, ith row and ith column of $\mathbf{X}$, respectively; $\mathbf{X}^\dagger$, $\mathbf{X}^H$ and $\mathbf{X}^\top$ mean the Moore-Penrose pseudo-inverse, conjugate transpose and transpose of $\mathbf{X}$, respectively; $\text{supp}(\mathbf{X})$ is the non-zero row indices (i.e., joint sparsity) of $\mathbf{X}$; $\|\mathbf{X}\|_F$ and $\|\mathbf{X}\|_2$ signify the Frobenius and Euclidean norm of $\mathbf{X}$, respectively. Moreover, S^c is the complement of set S; $\mathbf{I}_L$ is an $L\times L$ identity matrix.

I. Proof of Theorem 1

Theorem 1. *The actual sampling rate of (4) is* $\min\left(pf_s, f_{nyq}\right)$, *which attains the theoretical lower bound of sampling rate in MCS when* $|supp(\mathbf{X})| \le \frac{N_{sig}B}{f_s}$.

Proof. In the ith channel of a multi-coset sampler, the sampling sequence is given by

$$x_{c_i}[n] = x(LTn+\tau_i), \quad n = 0,1,\cdots . \tag{S.1}$$

The sampling rate of each channel is determined by the sampled signal sequence. To be specific, since the sampling time interval is LT, the sampling rate of each channel is

$$f_s = \frac{1}{LT} = \frac{f_{\text{nyq}}}{L}, \tag{S.2}$$

i.e., one-Lth of the Nyquist sampling rate.

Moreover, as the multi-coset sampler is assumed to have p channels, the overall sampling rate of p channels is p times that of each channel (i.e. $\frac{pf_{\text{nyq}}}{L}$). If this sampling rate is greater than the Nyquist rate f_{nyq}, then the advantage of sub-Nyquist sampling structure no longer exists. In this case, we only need to sample at Nyquist sampling rate f_{nyq}. Thus, the actual sampling rate can be given by

$$\min\left(\frac{p}{LT}, f_{\text{nyq}}\right). \tag{S.3}$$

The theoretical lower bound of the sampling rate is given in [17], which is determined directly by the true bandwidth of the signal:

$$\min(2\lambda(\mathcal{T}), f_{\text{nyq}}). \tag{S.4}$$

Thus, the theoretical lower bound on the sampling rate is achieved when

$$\min\left(\frac{p}{LT}, f_{\text{nyq}}\right) \le \min(2\lambda(\mathcal{T}), f_{\text{nyq}}). \tag{S.5}$$

In most cases, $2\lambda(\mathcal{T})$ and $\frac{p}{LT}$ do not exceed f_{nyq}. (If violated, the sampling rate would just be f_{nyq}.) Therefore, the condition (S.5) holds whenever

$$\frac{p}{LT} \le 2\lambda(\mathcal{T}). \tag{S.6}$$

Furthermore, to ensure an unique-solution reconstruction, the number p of channels should not be too small. In particular, it's lower bound is twice the signal sparsity without the priori information about the signal $\mathbf{X}$ [17],

$$p \ge 2|\text{supp}(\mathbf{X})|. \tag{S.7}$$

For the worst case where $p = 2|\text{supp}(\mathbf{X})|$, (S.6) can be rewritten as

$$|\text{supp}(\mathbf{X})| \le \lambda(\mathcal{T})LT = \frac{N_{\text{sig}}B}{f_s}, \tag{S.8}$$

which completes the proof. □

II. Proof of Theorem 2

Theorem 2. *When* $r \in [\lceil \frac{f_s}{\lfloor (D-1)f_s - B\rfloor}\rceil, N]$ *and* $B > f_s$, *we have* $\max_{j\in\{1,\cdots,r\}} \left|supp(\bar{\mathbf{X}}_{S_j})\right| < \frac{N_{sig}B}{f_s}$.

Proof. Recall that the MMV model $\mathbf{Y} = \mathbf{AX} + \mathbf{E}$ is decomposed into r sub-MMV problems:

$$\mathbf{Y}_{S_j} = \mathbf{A}\mathbf{X}_{S_j} + \mathbf{E}_{S_j}, \quad j = 1,\cdots,r \tag{S.9}$$

and each problem is solved individually. Our goal is to determine the number r of sub-MMV problems that ensures the actual sampling rate to reach the lower bound of the theoretical sampling rate.

For all row blocks $\{\mathbf{X}^{U_1},\cdots,\mathbf{X}^{U_M}\}$ in $\mathbf{X}$, assume that each block has consecutive frequency points of length B (i.e., the sub-band's width is B) if occupied and zero otherwise. In this work, we are primarily interested in the case where

$$B > f_s. \tag{S.10}$$

since the remaining case where $B \le f_s$ has been studied thoroughly in existing works (see. e.g., [16]).

Let D denote the height (i.e., number of rows) in each row block $\mathbf{X}^{U_i}$, which can be determined properly according to B and f_s. Specifically, we choose a $D \ge 3$ such that

$$(D-2)f_s < B \le (D-1)f_s. \tag{S.11}$$

In this case, the frequency points of each PU signal (i.e., occupied row block) occupy $D-1$ or D rows. An example of $D = 5$ is illustrated in Fig. 1, where PU signals 1 and 2 occupy 4 rows while PU signal 3 occupies 5 rows.

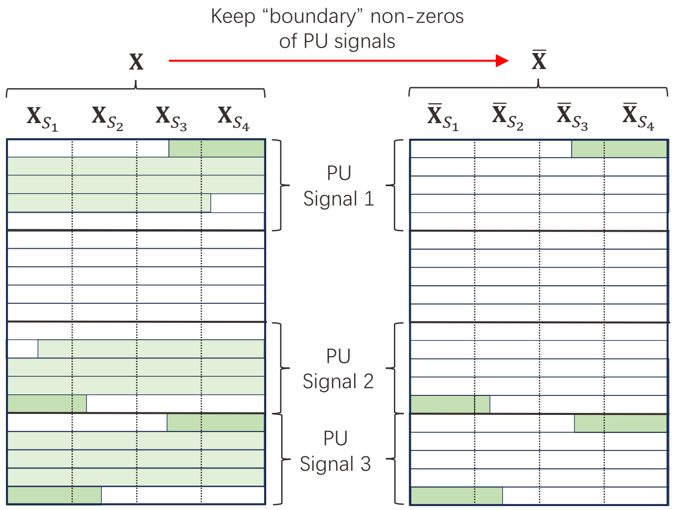

Fig. 1. An illustrative example of MCS signal $\mathbf{X}$ with 3 PU signals.

- When $D-1$ rows in PU signal $\mathbf{X}^{U_i}$ are occupied, only one row of $\bar{\mathbf{X}}^{U_i}$ is occupied, since $\bar{\mathbf{X}}^{U_i}$ only keeps the boundary non-zeros of $\mathbf{X}^{U_i}$. In this case, we can easily see that

$$\left|\text{supp}(\bar{\mathbf{X}}_{S_j}^{U_i})\right| \leq \left|\text{supp}(\bar{\mathbf{X}}^{U_i})\right| = 1. \tag{S.12}$$

Since there are N_{sig} sub-band signals in $\mathbf{X}$, and also noting that $B > f_s$, we have

$$\max_{j\in\{1,\cdots,r\}}\left|\text{supp}(\bar{\mathbf{X}}_{S_j})\right| \leq N_{\text{sig}} < \frac{N_{\text{sig}}B}{f_s}. \tag{S.13}$$

- When D rows in PU signal $\mathbf{X}^{U_i}$ are occupied, the length l of the frequency points in $\bar{\mathbf{X}}^{U_i}$ obeys

$$l = B - (D-2)f_s \overset{(\text{S.11})}{\leq} f_s. \tag{S.14}$$

As a result, the column indices of non-zeros in any PU signal $\bar{\mathbf{X}}^{U_i}$ do not overlap. In other words, the total length of frequency points in each PU signal $\bar{\mathbf{X}}^{U_i}$ does not exceed f_s, as shown in Fig. 1.

Recall that $\bar{\mathbf{X}} \in \mathbb{R}^{L\times N}$ is column-partitioned into r sub-matrices (i.e., $\{\bar{\mathbf{X}}_{S_j}\}_{j=1,\cdots,r}$).

 - We first consider an extreme case where $r = N$. In this case, $\bar{\mathbf{X}}_{S_j}$ only have one column (i.e., $|S_j| = |\{j\}| = 1$), and so each column-partitioned PU signal $\bar{\mathbf{X}}_{S_j}^{U_i}$ satisfies

$$\left|\text{supp}(\bar{\mathbf{X}}_{S_j}^{U_i})\right| = \left|\text{supp}(\bar{\mathbf{X}}_{j,:}^{U_i})\right| \leq 1. \tag{S.15}$$

Similar to (S.13), we further have

$$\begin{aligned}\max_{j\in\{1,\cdots,r\}}\left|\text{supp}(\bar{\mathbf{X}}_{S_j})\right| &\leq N_{\text{sig}} \\ &\overset{(\text{S.10})}{<} \frac{N_{\text{sig}}B}{f_s}.\end{aligned} \tag{S.16}$$

 - We then consider the general case where $r < N$. In this case, the length of frequency points in each sub-matrix $\bar{\mathbf{X}}_{S_j}^{U_i}$ does not exceed $\lceil \frac{f_s}{r} \rceil$. In the following, we shall prove that

$$\max_{j\in\{1,\cdots,r\}}\left|\text{supp}(\bar{\mathbf{X}}_{S_j})\right| \leq N_{\text{sig}} \tag{S.17}$$

when $r \in [\lceil \frac{f_s}{\lfloor (D-1)f_s - B\rfloor} \rceil, N)$.

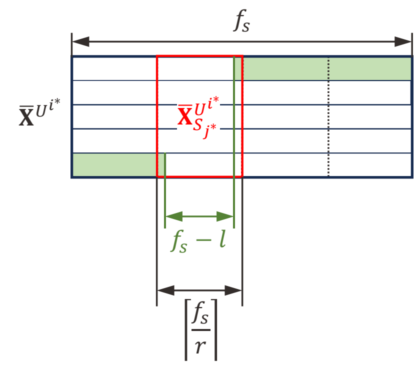

Fig. 2. An illustrative example of the PU signal $\bar{\mathbf{X}}^{U_{i^*}}$, which has a length l that is greater than $f_s - \lceil \frac{f_s}{r} \rceil$.

Assume that there exists some a sub-matrix, say, $\bar{\mathbf{X}}_{S_{j^*}}$, such that

$$\left|\text{supp}(\bar{\mathbf{X}}_{S_{j^*}})\right| > N_{\text{sig}}. \tag{S.18}$$

Then, there exists at least a row-partitioned block of $\bar{\mathbf{X}}_{S_{j^*}}$, say, $\bar{\mathbf{X}}_{S_{j^*}}^{U_{i^*}}$, that satisfies

$$\left|\text{supp}(\bar{\mathbf{X}}_{S_{j^*}}^{U_{i^*}})\right| = 2. \tag{S.19}$$

This implies that the length l of the frequency points in PU signal $\bar{\mathbf{X}}^{U_{i^*}}$ must be greater than $f_s - \lceil \frac{f_s}{r} \rceil$, as illustrated in Fig. 2. That is,

$$\left\lceil \frac{f_s}{r} \right\rceil > f_s - l, \tag{S.20}$$

which, together with the fact that $l = B - (D-2)f_s$ (see (S.14)), leads to

$$\begin{aligned}\left\lceil \frac{f_s}{r} \right\rceil &> (D-1)f_s - B \\ &> \lfloor (D-1)f_s - B \rfloor.\end{aligned} \tag{S.21}$$

Since $\lfloor (D-1)f_s - B \rfloor$ is a integer, we have

$$r < \frac{f_s}{\lfloor (D-1)f_s - B \rfloor}. \tag{S.22}$$

Thus, if

$$r \geq \frac{f_s}{\lfloor (D-1)f_s - B \rfloor}, \tag{S.23}$$

the assumption of (S.18) must not be true, which implies (S.17).

To sum up, when $r \in [\lceil \frac{f_s}{\lfloor (D-1)f_s - B\rfloor} \rceil, N]$, we have

$$\max_{j\in\{1,\cdots,r\}}\left|\text{supp}(\bar{\mathbf{X}}_{S_j})\right| \leq N_{\text{sig}} < \frac{N_{\text{sig}}B}{f_s}. \tag{S.24}$$

The proof is thus complete. □

III. Proof of Theorem 3

Theorem 3. *Consider the column-partitioned MMV model (5) with* $\min_{i,j}\|(\mathbf{X}_{S_i})_{j,:}\|_2/\|\mathbf{X}_{S_i}\|_F = \eta$ *and* $|supp(\mathbf{X}_{S_i})| \leq s$. *Let* $s_1 := \min_{i,k}|\Lambda_{S_i}^k \cap supp(\mathbf{X}_{S_i})|$, $s_2 := \min_{i,k}|\Lambda_{S_i}^k \cap$

$supp(\mathbf{X}_{S_i}) \cap \tilde{S}^k_{S_i}|$ *and* $s_3 := \min_{i,k}|\Lambda^k_{S_i} \cap supp(\mathbf{X}_{S_i}) \cap \tilde{S}^k_{S_i} \setminus S^k_{S_i}|$. *Then, if the sensing matrix* $\mathbf{A}$ *obeys the RIP with*

$$\delta_{3s} \le \sqrt{\frac{\nu_1\sqrt{\nu_1^2+4\nu_2^2}-\nu_1^2-1}{4\nu_1^2\nu_2^2-2\nu_1^2-1}} \tag{S.25}$$

where $\nu_1 := \frac{1+\omega}{1+\eta\omega\sqrt{s_2}}$ *and* $\nu_2 := \frac{1+\omega}{1+\eta\omega\sqrt{s_3}}$, *SI-SSP produces an signal estimate* $\mathbf{X}^k = [\mathbf{X}^k_{S_1}, \cdots, \mathbf{X}^k_{S_r}]$ *satisfying*

$$\left\|\mathbf{X}-\mathbf{X}^k\right\|_F \le \rho^k \left\|\mathbf{X}\right\|_F + \tau \left\|\mathbf{E}\right\|_F, \tag{S.26}$$

where $\rho \in (0,1)$ *and* τ *are constants depending on* δ_{3s}, ν_1 *and* ν_2. *Furthermore, after at most* $k^* = \lceil \log_\rho \frac{\|\mathbf{X}\|_F}{\tau\|\mathbf{E}\|_F} \rceil$ *iterations, SI-SSP estimates* $\mathbf{X}$ *with*

$$\|\mathbf{X}-\mathbf{X}^{k^*}\|_F \le (\tau+1)\|\mathbf{E}\|_F. \tag{S.27}$$

To prove Theorem 3, we first introduce six useful Lemmas, whose proofs are left to the appendices.

Lemma 1. *([25]): For nonnegative numbers a, b, c, d, x, y,*

$$(ax+by)^2+(cx+dy)^2 \le (\sqrt{a^2+c^2}x+(b+d)y)^2. \tag{S.28}$$

Lemma 2. *Consider the system model* $\mathbf{Y} = \mathbf{AX}+\mathbf{E}$, *where* $supp(\mathbf{X}) = T$ *and* $|T| = s$. *Let* $S \subseteq \{1,2,...,n\}$ *be an index set with* $|S| = t$ *and* $\mathbf{W}_{T_0}$ *be a side-information matrix with diagonal entries indexed by* $T_0 \subseteq \{1,2,...,n\}$ *being* $\omega \ge 0$ *and zero otherwise. Also, let* $\tilde{\mathbf{X}} := \arg\min_{\mathbf{Z}:supp(\mathbf{Z})\subseteq S} \|\mathbf{Y}-\mathbf{AZ}\|_2$. *If* $\delta_{3s} < 1$, *then*

$$\|\mathbf{W}_{T_0}(\mathbf{X}-\tilde{\mathbf{X}})_S\|_F \le \omega\delta_{s+t}\|\mathbf{X}-\tilde{\mathbf{X}}\|_F + \omega\sqrt{1+\delta_t}\|\mathbf{E}\|_F \tag{S.29}$$

and

$$\|\mathbf{X}-\tilde{\mathbf{X}}\|_F \le \sqrt{\frac{1}{1-\delta_{s+t}^2}}\|\mathbf{X}_{S^c}\|_F + \frac{\sqrt{1+\delta_t}}{1-\delta_{s+t}}\|\mathbf{E}\|_F. \tag{S.30}$$

Furthermore, if $t > s$, *define* $T_\triangledown$ *as the row-indices of the smallest* $t-s$ *row-norm entries of* $\tilde{\mathbf{X}}$ *in* S, *we have*

$$\|\mathbf{X}_{T_\triangledown}\|_F \le \sqrt{2}\nu_2\delta_{s+t}\|\mathbf{X}-\tilde{\mathbf{X}}\|_F + \nu_2\sqrt{2\left(1+\delta_t\right)}\|\mathbf{E}\|_F. \tag{S.31}$$

Remark 1. *When we consider the atom selection strategy of* $\|\tilde{\mathbf{X}}_{T_\triangledown} + \mathbf{W}_{T_0}\tilde{\mathbf{X}}_{T_\triangledown}\|_F \le \|\tilde{\mathbf{X}}_{S'} + \mathbf{W}_{T_0}\tilde{\mathbf{X}}_{S'}\|_F$, *we can also obtain another upper bound for* $\|\mathbf{X}_{T_\triangledown}\|_F$ *in* (S.31). *In this case, we should allocate* $2\left\|\mathbf{X}_{T_\triangledown}\right\|_F$ *to the left hand side of* (A.55), *we have*

$$\|\mathbf{X}_{T_\triangledown}\|_F \le \sqrt{2}\nu_3\delta_{s+t}\|\mathbf{X}-\tilde{\mathbf{X}}\|_F + \nu_4\sqrt{2\left(1+\delta_t\right)}\left\|\mathbf{E}\right\|_F. \tag{S.32}$$

where $\nu_3 = (1-\omega+\omega\delta_{s+t}+\delta_{s+t})/(2\delta_{s+t})$ *and* $\nu_4 = (1+\omega)/(2\delta_{s+t})$.

Lemma 3. *In steps 4 and 5 of SI-SSP, we have*

$$\|\mathbf{X}_{(\tilde{S}^k)^c}\|_F \le \sqrt{2}\nu_1\delta_{3s}\left\|\mathbf{X}-\mathbf{X}^{k-1}\right\|_F + \nu_1\sqrt{2(1+\delta_{3s})}\left\|\mathbf{E}\right\|_F. \tag{S.33}$$

Remark 2. *When we consider the atom selection strategy in select step that*

$$\begin{aligned}&\left\|((\mathbf{I}_L+\mathbf{W}_{T_0})\mathbf{A}^H(\mathbf{Y}-\mathbf{AX}^{k-1}))_T\right\|_F\\ &\quad\le \left\|((\mathbf{I}_L+\mathbf{W}_{T_0})\mathbf{A}^H(\mathbf{Y}-\mathbf{AX}^{k-1}))_{\Delta S}\right\|_F.\end{aligned} \tag{S.34}$$

We can also obtain another upper bound for $\|\mathbf{X}_{(\tilde{S}^k)^c}\|_F$ *in* (S.33). *In this case, we should allocate* $2\|\mathbf{X}_{(\tilde{S}^k)^c}\|_F$ *to the left hand side of* (A.68), *we have*

$$\begin{aligned}\|\mathbf{X}_{(\tilde{S}^k)^c}\|_F \le{}& \sqrt{2}\nu_3\delta_{3s}\left\|\mathbf{X}-\mathbf{X}^{k-1}\right\|_F\\ &+\nu_4\sqrt{2\left(1+\delta_{3s}\right)}\left\|\mathbf{E}\right\|_F.\end{aligned} \tag{S.35}$$

where $\nu_3 = (1-\omega+\omega\delta_{3s}+\delta_{3s})/(2\delta_{3s})$ *and* $\nu_4 = (1+\omega)/(2\delta_{3s})$. *Based on conclusions* (S.32) *and* (S.35), *we know that the sensing matrix* $\mathbf{A}$ *obeys the RIP with*

$$\delta_{3s} \le \sqrt{\frac{\nu_3\sqrt{\nu_3^2+4\nu_4^2}-\nu_3^2-1}{4\nu_3^2\nu_4^2-2\nu_3^2-1}}. \tag{S.36}$$

Lemma 4. *Let* $T_0 \subseteq \{1,2,...,n\}$, *for two vectors* $\mathbf{u},\mathbf{v} \in \mathbb{R}^n$, *if* $|supp(\mathbf{u}) \cup supp(\mathbf{v})| \le t$,

$$\left|\left\langle \mathbf{u}, (\mathbf{W}_{T_0}-\mathbf{W}_{T_0}\mathbf{A}^H\mathbf{A})\mathbf{v}\right\rangle\right| \le \omega\delta_t\left\|\mathbf{u}\right\|\left\|\mathbf{v}\right\|; \tag{S.37}$$

Moreover, if $U \subseteq \{1,2,...,n\}$ *and* $|U \cup supp(\mathbf{v})| \le t$, *then*

$$\left|(\mathbf{W}_{T_0}-\mathbf{W}_{T_0}\mathbf{A}^H\mathbf{A})\mathbf{v}\right| \le \omega\delta_t\left\|\mathbf{v}\right\|. \tag{S.38}$$

Lemma 5. *For SMV model* $\mathbf{y} = \mathbf{\Phi x}+\mathbf{e}$, *let* $T_0 \subseteq \{1,2,...,n\}$, *let* $U \subseteq \{1,2,...,n\}$ *and* $|U \cap T_0| \le u$, *we have*

$$\left\|(\mathbf{W}_{T_0}\mathbf{A}^H\mathbf{e})_U\right\|_2 \le \omega\delta_u\left\|\mathbf{e}\right\|_2. \tag{S.39}$$

Lemma 6. *Consider the MMV model* $\mathbf{Y} = \mathbf{AX}+\mathbf{E}$, *let* $\tilde{\mathbf{X}}$ *be the solution of the least squares problem* $\arg\min_{\mathbf{Z}}\{\|\mathbf{Y}-\mathbf{AZ}\|_F, supp(\mathbf{Z}) \subseteq S\}$, *then*

$$\langle \mathbf{W}_{T_0}\mathbf{X}-\mathbf{W}_{T_0}\tilde{\mathbf{X}}, \mathbf{A}^H\mathbf{AZ}\rangle + \omega\left\langle \mathbf{E}, \mathbf{AZ}\right\rangle = \mathbf{0}. \tag{S.40}$$

Now we have all ingredients to prove Theorem 3.

Proof of Theorem 3. First, in Steps 4 and 5 of SI-SSP, Lemma 3 implies

$$\|\mathbf{X}_{(\tilde{S}^k)^c}\|_F \le \sqrt{2}\nu_1\delta_{3s}\left\|\mathbf{X}-\mathbf{X}^{k-1}\right\|_F + \nu_1\sqrt{2(1+\delta_{3s})}\left\|\mathbf{E}\right\|_F. \tag{S.41}$$

Note that Step 6 of SI-SSP solves a least squares problem. Let $S = \tilde{S}^k$ and $\tilde{\mathbf{X}} = \tilde{\mathbf{X}}^k$, $t = 2s$, by (S.30) we have

$$\|\mathbf{X}-\tilde{\mathbf{X}}\|_F \le \sqrt{\frac{1}{1-\delta_{3s}^2}}\|\mathbf{X}_{(\tilde{S}^k)^c}\|_F + \frac{\sqrt{1+\delta_{2s}}}{1-\delta_{3s}}\|\mathbf{E}\|_F. \tag{S.42}$$

Combining (S.41) and (S.42) and also magnifying δ_{2s} to δ_{3s}, we further have

$$\|\mathbf{X}-\tilde{\mathbf{X}}^k\|_F \le \nu_1\sqrt{\frac{2\delta_{3s}^2}{1-\delta_{3s}^2}}\|\mathbf{X}-\mathbf{X}^{k-1}\|_F + \tau_1\|\mathbf{E}\|_F. \tag{S.43}$$

Next, after Step 7 of SI-SSP, let $S_\triangledown = \tilde{S}^k \setminus S^k$ be the row-indices of the smallest $t-s$ row norm entries in $\tilde{\mathbf{X}}^k$. Also, let $T = \tilde{S}^k$, $\tilde{\mathbf{X}} = \tilde{\mathbf{X}}^k$, $T_\triangledown = S_\triangledown$ and $t = 2s$. Then, by (A.54) we have

$$\|\mathbf{X}_{S_\triangledown}\|_F \le \sqrt{2}\nu_2\delta_{3s}\|\mathbf{X}-\tilde{\mathbf{X}}^k\|_F + \nu_2\sqrt{2\left(1+\delta_{2s}\right)}\|\mathbf{E}\|_F. \tag{S.44}$$

Let $\tau_1 = (\nu_1\sqrt{2(1-\delta_{3s})}+\sqrt{1+\delta_{3s}})(1-\delta_{3s})^{-1}$ and $\tau_2 = \sqrt{1+\delta_{3s}}$. Dividing $(S^k)^c$ into two disjoint subsets: $(\tilde{S}^k)^c$ and $S_\triangledown$, we get

$$\left\|\mathbf{X}_{(S^k)^c}\right\|_F^2 = \|\mathbf{X}_{S_\triangledown}\|_F^2 + \|\mathbf{X}_{(\tilde{S}^k)^c}\|_F^2$$

$$\overset{\text{(S.41), (S.44)}}{\leq} 2\left(\nu_2\delta_{3s}\|\mathbf{X}-\tilde{\mathbf{X}}^k\|_F+\nu_2\tau_2\|\mathbf{E}\|_F\right)^2 + 2\left(\nu_1\delta_{3s}\left\|\mathbf{X}-\mathbf{X}^{k-1}\right\|_2+\nu_1\tau_2\left\|\mathbf{E}\right\|_F\right)^2$$
$$\overset{\text{(S.43)}}{\leq} 2\left(\sqrt{\frac{2\nu_1^2\nu_2^2\delta_{3s}^4}{1-\delta_{3s}^2}}\left\|\mathbf{X}-\mathbf{X}^{k-1}\right\|_F+\nu_2(\tau_1\delta_{3s}+\tau_2) \times\|\mathbf{E}\|_F\right)^2+2\left(\nu_1\delta_{3s}\left\|\mathbf{X}-\mathbf{X}^{k-1}\right\|_F+\nu_1\tau_2\left\|\mathbf{E}\right\|_F\right)^2$$
$$\overset{\text{(S.28)}}{\leq} 2\left(\sqrt{\frac{2\nu_1^2\nu_2^2\delta_{3s}^4}{1-\delta_{3s}^2}+\nu_1^2\delta_{3s}^2}\left\|\mathbf{X}-\mathbf{X}^{k-1}\right\|_F + \left((\nu_1+\nu_2)\tau_2+\nu_2\delta_{3s}\tau_1\right)\left\|\mathbf{E}\right\|_F\right)^2. \quad \text{(S.45)}$$

Squaring both sides, we get

$$\left\|\mathbf{X}_{(S^k)^c}\right\|_F \leq \sqrt{\tfrac{4\nu_1^2\nu_2^2\delta_{3s}^4}{1-\delta_{3s}^2}+2\nu_1^2\delta_{3s}^2}\left\|\mathbf{X}-\mathbf{X}^{k-1}\right\|_F + \sqrt{2}\left((\nu_1+\nu_2)\tau_2+\nu_2\delta_{3s}\tau_1\right)\left\|\mathbf{E}\right\|_F. \quad \text{(S.46)}$$

Step 9 of SI-SSP also solves a least squares problem. Letting $T=S^k$, $\tilde{\mathbf{X}}=\mathbf{X}^k$ and $t=s$, by (S.30), we have

$$\|\mathbf{X}-\mathbf{X}^k\|_F \leq \sqrt{\frac{1}{1-\delta_{2s}^2}}\|\mathbf{X}_{(S^k)^c}\|_F+\frac{\sqrt{1+\delta_s}}{1-\delta_{2s}}\|\mathbf{E}\|_F. \quad \text{(S.47)}$$

Finally, combining (S.46) and (S.47) yields

$$\|\mathbf{X}-\mathbf{X}^k\|_F \leq \rho\|\mathbf{X}-\mathbf{X}^{k-1}\|_F+(1-\rho)\tau\|\mathbf{E}\|_F \quad \text{(S.48)}$$

where $\rho := \sqrt{2}\delta_{3s}\sqrt{2\nu_1^2\nu_2^2\delta_{3s}^2+\nu_1^2-\nu_1^2\delta_{3s}^2}(1-\delta_{3s}^2)^{-1}$ and $\tau := \sqrt{2}\delta_{3s}\nu_2(\nu_1\sqrt{2(1-\delta_{3s})}+\sqrt{1+\delta_{3s}})(1-\delta_{3s}^2)^{-1/2}(1-\delta_{3s})^{-1}(1-\rho)^{-1}+(\nu_1\nu_2\sqrt{2(1-\delta_{3s})}+\sqrt{1+\delta_{3s}})(1-\delta_{3s})^{-1}$.

We recursively apply (S.48) to obtain

$$\|\mathbf{X}-\mathbf{X}^k\|_F \leq \rho^k\|\mathbf{X}\|_F+\tau\|\mathbf{E}\|_F \quad \text{(S.49)}$$

where $\rho<1$ under (S.25). When $k^*=\lceil\log_\rho\frac{\|\mathbf{X}\|_F}{\tau\|\mathbf{E}\|_F}\rceil$, we have $\rho^k\|\mathbf{X}\|_F\leq\tau\|\mathbf{E}\|_F$, and thus the stability result (S.27). $\square$

Appendix A
Proof of Lemma 2

- First, we give a upper bound of $\|\mathbf{X}_{T_\nabla}\|_F$, by Lemma 6, let $\mathbf{Z}=(\mathbf{W}_{T_0}\mathbf{X}-\mathbf{W}_{T_0}\tilde{\mathbf{X}})_S$, we have

$$\left\langle\mathbf{W}_{T_0}(\mathbf{X}-\tilde{\mathbf{X}}),\mathbf{A}^H\mathbf{A}(\mathbf{W}_{T_0}\mathbf{X}-\mathbf{W}_{T_0}\tilde{\mathbf{X}})_S\right\rangle + \left\langle\mathbf{W}_{T_0}\mathbf{E},\mathbf{A}(\mathbf{W}_{T_0}\mathbf{X}-\mathbf{W}_{T_0}\tilde{\mathbf{X}})_S\right\rangle=\mathbf{0}. \quad \text{(A.50)}$$

Noticing that supp$(\tilde{\mathbf{X}})\subseteq S$, we have

$$\begin{aligned}
&\|(\mathbf{W}_{T_0}\mathbf{X}-\mathbf{W}_{T_0}\tilde{\mathbf{X}})_S\|_F^2\\
=&\left\langle\mathbf{W}_{T_0}(\mathbf{X}-\tilde{\mathbf{X}}),(\mathbf{W}_{T_0}\mathbf{X}-\mathbf{W}_{T_0}\tilde{\mathbf{X}})_S\right\rangle\\
\overset{\text{(A.50)}}{=}&\left\langle\mathbf{W}_{T_0}(\mathbf{X}-\tilde{\mathbf{X}}),(\mathbf{I}_L-\mathbf{A}^H\mathbf{A})(\mathbf{X}-\tilde{\mathbf{X}})_S\right\rangle\\
-&\left\langle\mathbf{W}_{T_0}\mathbf{E},\mathbf{A}(\mathbf{W}_{T_0}\mathbf{X}-\mathbf{W}_{T_0}\tilde{\mathbf{X}})_S\right\rangle\\
\overset{\text{(7)}}{\leq}&\ \omega\delta_{s+t}\left\|(\mathbf{W}_{T_0}\mathbf{X}-\mathbf{W}_{T_0}\tilde{\mathbf{X}})_S\right\|_F\left\|\mathbf{X}-\tilde{\mathbf{X}}\right\|_F\\
&+\omega\left\|\mathbf{E}\right\|_F\sqrt{1+\delta_t}\left\|\mathbf{W}_{T_0}(\mathbf{X}-\tilde{\mathbf{X}})_S\right\|_F. \quad \text{(A.51)}
\end{aligned}$$

Divide both sides by $\|(\mathbf{W}_{T_0}\mathbf{X}-\mathbf{W}_{T_0}\tilde{\mathbf{X}})_S\|_F$ to obtain (S.29).

- Next, by expanding [Lemma 2, 25] to the MMV model, we could get a relationship between $\|\mathbf{X}-\tilde{\mathbf{X}}\|_F$ and $\|\mathbf{X}_{S^c}\|_F$. We have

$$\|\mathbf{X}-\tilde{\mathbf{X}}\|_F \leq \sqrt{\frac{1}{1-\delta_{s+t}^2}\|\mathbf{X}_{S^c}\|_F^2}+\frac{\sqrt{1+\delta_t}}{1-\delta_{s+t}}\|\mathbf{E}\|_F. \quad \text{(A.52)}$$

- Then, we established the relationship between $\mathbf{X}_{T_\nabla}$ and $\mathbf{X}-\tilde{\mathbf{X}}$. There exist a subset $S'\subseteq S$ and $S'\cap T=\emptyset$. Since T_∇ is defined by the set of indices of the $t-s$ smallest row entries of $\tilde{\mathbf{X}}$, we can conclude that

$$\begin{aligned}
&\|\tilde{\mathbf{X}}_{T_\nabla}\|_F+\|\mathbf{W}_{T_0}\tilde{\mathbf{X}}_{T_\nabla}\|_F\\
\leq\ &\|\tilde{\mathbf{X}}_{S'}\|_F+\|\mathbf{W}_{T_0}\tilde{\mathbf{X}}_{S'}\|_F. \quad \text{(A.53)}
\end{aligned}$$

By eliminating the contribution from $T_\nabla\cap S'$ and noticing that $S'\cap T=\emptyset$, we have

$$\begin{aligned}
\|\tilde{\mathbf{X}}_{T_\nabla\setminus S'}\|_F\ &+\ \|\mathbf{W}_{T_0}\tilde{\mathbf{X}}_{T_\nabla\setminus S'}\|_F\\
&\leq\ \|(\tilde{\mathbf{X}}-\mathbf{X})_{S'\setminus T_\nabla}\|_F\\
&+\ \|\mathbf{W}_{T_0}(\tilde{\mathbf{X}}-\mathbf{X})_{S'\setminus T_\nabla}\|_F. \quad \text{(A.54)}
\end{aligned}$$

For the left-hand side of (A.54), we have

$$\begin{aligned}
&\|\tilde{\mathbf{X}}_{T_\nabla\setminus S'}\|_F+\|\mathbf{W}_{T_0}\tilde{\mathbf{X}}_{T_\nabla\setminus S'}\|_F\\
=\ &\|(\tilde{\mathbf{X}}-\mathbf{X}+\mathbf{X})_{T_\nabla\setminus S'}\|_F\\
+\ &\|(\mathbf{W}_{T_0}(\tilde{\mathbf{X}}-\mathbf{X})+\mathbf{W}_{T_0}\mathbf{X})_{T_\nabla\setminus S'}\|_F\\
\geq\ &\|\mathbf{X}_{T_\nabla}\|_F+\|\mathbf{W}_{T_0}\mathbf{X}_{T_\nabla}\|_F \quad \text{(A.55)}\\
-\ &\|(\tilde{\mathbf{X}}-\mathbf{X})_{T_\nabla\setminus S'}\|_F\\
-\ &\|\mathbf{W}_{T_0}(\tilde{\mathbf{X}}-\mathbf{X})_{T_\nabla\setminus S'}\|_F. \quad \text{(A.56)}
\end{aligned}$$

Finally, combining (A.56) and (A.54), and noticing that

$$(T_\nabla\setminus S')\cap(S'\setminus T_\nabla)\ =\ \emptyset \quad \text{(A.57)}$$
$$(T_\nabla\setminus S')\cup(S'\setminus T_\nabla)\ \subseteq\ T, \quad \text{(A.58)}$$

we have

$$\begin{aligned}
&\|\mathbf{X}_{T_\nabla}\|_F+\|\mathbf{W}_{T_0}\mathbf{X}_{T_\nabla}\|_F\\
\leq\ &\|(\tilde{\mathbf{X}}-\mathbf{X})_{T_\nabla\setminus S'}\|_F+\|(\mathbf{W}_{T_0}(\tilde{\mathbf{X}}-\mathbf{X}))_{T_\nabla\setminus S'}\|_F\\
+\ &\|(\tilde{\mathbf{X}}-\mathbf{X})_{S'\setminus T_\nabla}\|_F+\|(\mathbf{W}_{T_0}(\tilde{\mathbf{X}}-\mathbf{X}))_{S'\setminus T_\nabla}\|_F\\
\leq\ &\sqrt{2}\|(\tilde{\mathbf{X}}-\mathbf{X})_{(S'\setminus T_\nabla)\cup(T_\nabla\setminus S')}\|_F\\
+\ &\sqrt{2}\|(\mathbf{W}_{T_0}(\tilde{\mathbf{X}}-\mathbf{X}))_{(S'\setminus T_\nabla)\cup(T_\nabla\setminus S')}\|_F\\
\leq\ &\sqrt{2}\|(\tilde{\mathbf{X}}-\mathbf{X})_S\|_F+\sqrt{2}\|(\mathbf{W}_{T_0}(\tilde{\mathbf{X}}-\mathbf{X}))_S\|_F\\
\overset{\text{(S.29)}}{\leq}\ &\sqrt{2}(1+\omega)\delta_{s+t}\|\mathbf{X}-\tilde{\mathbf{X}}\|_F\\
+\ &(1+\omega)\sqrt{2(1+\delta_t)}\|\mathbf{E}\|_F. \quad \text{(A.59)}
\end{aligned}$$

Also, we can obtain the relationship between $\|\mathbf{W}_{T_0}\mathbf{X}_{T_\nabla}\|_F$ and $\|\mathbf{X}_{T_\nabla}\|_F$:

$$\eta\omega\sqrt{s_3}\left\|\mathbf{X}_{T_\nabla}\right\|_F\leq\left\|\mathbf{W}_{T_0}\mathbf{X}_{T_\nabla}\right\|_F. \quad \text{(A.60)}$$

Combining (A.59) and (A.60), we have

$$\begin{aligned}
\|\mathbf{X}_{T_\nabla}\|_F\ \leq\ &\frac{\sqrt{2}(1+\omega)\delta_{s+t}}{1+\eta\omega\sqrt{s_3}}\|\mathbf{X}-\tilde{\mathbf{X}}\|_F\\
+\ &\frac{(1+\omega)\sqrt{2(1+\delta_t)}}{1+\eta\omega\sqrt{s_3}}\left\|\mathbf{E}\right\|_F. \quad \text{(A.61)}
\end{aligned}$$

Noting the definition of ν_2, we complete the proof of Lemma 2.

Appendix B
Proof of Lemma 3

Proof: From Step 5 of SI-SSP, we have

$$\mathbf{X}_{S_i}^k = \underset{\boldsymbol{\Theta}:\text{supp}(\boldsymbol{\Theta})=S_{S_i}^k}{\arg\min} \|\mathbf{Y}_{S_i} - \mathbf{A}\boldsymbol{\Theta}\|_F. \tag{A.62}$$

From Step 4 of SI-SSP, let $\mathbf{X}^k = [\mathbf{X}_{S_1}^k, \cdots, \mathbf{X}_{S_r}^k]$. We have the following conclusion

$$\begin{aligned} &\|(\mathbf{A}^H(\mathbf{Y} - \mathbf{A}\mathbf{X}^{k-1}))_T\|_F \\ &+ \|(\mathbf{W}_{T_0}\mathbf{A}^H(\mathbf{Y} - \mathbf{A}\mathbf{X}^{k-1}))_T\|_F \\ &\leq \|(\mathbf{A}^H(\mathbf{Y} - \mathbf{A}\mathbf{X}^{k-1}))_{\Delta S}\|_F \\ &+ \|(\mathbf{W}_{T_0}\mathbf{A}^H(\mathbf{Y} - \mathbf{A}\mathbf{X}^{k-1}))_{\Delta S}\|_F. \end{aligned} \tag{A.63}$$

By removing the same coordinates $T \cap \Delta S$, we get

$$\begin{aligned} &\|(\mathbf{A}^H(\mathbf{Y} - \mathbf{A}\mathbf{X}^{k-1}))_{T\setminus\Delta S}\|_F \\ &+ \|(\mathbf{W}_{T_0}\mathbf{A}^H(\mathbf{Y} - \mathbf{A}\mathbf{X}^{k-1}))_{T\setminus\Delta S}\|_F \\ &\leq \|(\mathbf{A}^H(\mathbf{Y} - \mathbf{A}\mathbf{X}^{k-1}))_{\Delta S\setminus T}\|_F \\ &+ \|(\mathbf{W}_{T_0}\mathbf{A}^H(\mathbf{Y} - \mathbf{A}\mathbf{X}^{k-1}))_{\Delta S\setminus T}\|_F. \end{aligned} \tag{A.64}$$

Because $\text{supp}(\mathbf{X}) = T$ and $\text{supp}(\mathbf{X}^{k-1}) = S^{k-1}$,

$$(\mathbf{X} - \mathbf{X}^{k-1})_{\Delta S\setminus(T\cup S^{k-1})} = 0. \tag{A.65}$$

For the right-hand side of (A.64), we have

$$\begin{aligned} &\|(\mathbf{A}^H(\mathbf{Y} - \mathbf{A}\mathbf{X}^{k-1}))_{\Delta S\setminus T}\|_F \\ &+ \|(\mathbf{W}_{T_0}\mathbf{A}^H(\mathbf{Y} - \mathbf{A}\mathbf{X}^{k-1}))_{\Delta S\setminus T}\|_F \\ &= \|(\mathbf{A}^H(\mathbf{Y} - \mathbf{A}\mathbf{X}^{k-1}))_{\Delta S\setminus(T\cup S^{k-1})}\|_F \\ &+ \|(\mathbf{W}_{T_0}\mathbf{A}^H(\mathbf{Y} - \mathbf{A}\mathbf{X}^{k-1}))_{\Delta S\setminus(T\cup S^{k-1})}\|_F \\ &= \|(\mathbf{A}^H(\mathbf{A}\mathbf{X} + \mathbf{E} - \mathbf{A}\mathbf{X}^{k-1}))_{\Delta S\setminus(T\cup S^{k-1})}\|_F \\ &+ \|(\mathbf{W}_{T_0}\mathbf{A}^H(\mathbf{A}\mathbf{X} + \mathbf{E} - \mathbf{A}\mathbf{X}^{k-1}))_{\Delta S\setminus(T\cup S^{k-1})}\|_F \\ &\leq \|((\mathbf{A}^H\mathbf{A} - \mathbf{I}_L)(\mathbf{X} - \mathbf{X}^{k-1}))_{\Delta S\setminus T}\|_F \\ &+ \|(\mathbf{A}^H\mathbf{E})_{\Delta S\setminus T}\|_F \\ &+ \|(\mathbf{W}_{T_0}(\mathbf{A}^H\mathbf{A} - \mathbf{I}_L)(\mathbf{X} - \mathbf{X}^{k-1}))_{\Delta S\setminus T}\|_F \\ &+ \|(\mathbf{W}_{T_0}\mathbf{A}^H\mathbf{E})_{\Delta S\setminus T}\|_F. \end{aligned} \tag{A.66}$$

Note that $\tilde{S}^k = S^{k-1} \cup \Delta S$, we have

$$(\mathbf{X} - \mathbf{X}^{k-1})_{T\setminus\tilde{S}^k} = \mathbf{X}_{(\tilde{S}^k)^c}. \tag{A.67}$$

For the left-hand side of (A.64), we have

$$\begin{aligned} &\|(\mathbf{A}^H(\mathbf{Y} - \mathbf{A}\mathbf{X}^{k-1}))_{T\setminus\Delta S}\|_F \\ &+ \|(\mathbf{W}_{T_0}\mathbf{A}^H(\mathbf{Y} - \mathbf{A}\mathbf{X}^{k-1}))_{T\setminus\Delta S}\|_F \\ &= \|(\mathbf{A}^H(\mathbf{Y} - \mathbf{A}\mathbf{X}^{k-1}))_{T\setminus(\Delta S\cup S^{k-1})}\|_F \\ &+ \|(\mathbf{W}_{T_0}\mathbf{A}^H(\mathbf{Y} - \mathbf{A}\mathbf{X}^{k-1}))_{T\setminus(\Delta S\cup S^{k-1})}\|_F \\ &= \|(\mathbf{A}^H(\mathbf{A}\mathbf{X} + \mathbf{E} - \mathbf{A}\mathbf{X}^{k-1}))_{T\setminus(\Delta S\cup S^{k-1})}\|_F \\ &+ \|(\mathbf{W}_{T_0}\mathbf{A}^H(\mathbf{A}\mathbf{X} + \mathbf{E} - \mathbf{A}\mathbf{X}^{k-1}))_{T\setminus(\Delta S\cup S^{k-1})}\|_F \\ &= \|((\mathbf{A}^H\mathbf{A} - \mathbf{I}_L)(\mathbf{X} - \mathbf{X}^{k-1}) \\ &+ \mathbf{A}^H\mathbf{E} + \mathbf{X})_{\Delta S\setminus(T\cup S^{k-1})}\| \\ &+ \|\mathbf{W}_{T_0}(\mathbf{A}^H\mathbf{A} - \mathbf{I}_L)(\mathbf{X} - \mathbf{X}^{k-1}) \\ &+ \mathbf{W}_{T_0}\mathbf{A}^H\mathbf{E} + \mathbf{W}_{T_0}\mathbf{X})_{\Delta S\setminus(T\cup S^{k-1})}\|_F \\ &\geq \|\mathbf{X}_{(\tilde{S}^k)^c}\|_F + \|(\mathbf{W}_{T_0}\mathbf{X})_{(\tilde{S}^k)^c}\|_F \\ &- \|((\mathbf{A}^H\mathbf{A} - \mathbf{I}_L)(\mathbf{X} - \mathbf{X}^{k-1}))_{(\tilde{S}^k)^c}\|_F \end{aligned} \tag{A.68}$$

$$\begin{aligned} &- \|(\mathbf{W}_{T_0}(\mathbf{A}^H\mathbf{A} - \mathbf{I}_L)(\mathbf{X} - \mathbf{X}^{k-1}))_{(\tilde{S}^k)^c}\|_F \\ &- \|(\mathbf{A}^H\mathbf{E})_{(\tilde{S}^k)^c}\|_F - \|(\mathbf{W}_{T_0}\mathbf{A}^H\mathbf{E})_{(\tilde{S}^k)^c}\|_F. \end{aligned} \tag{A.69}$$

Combining (A.70) and (A.69), we have

$$\begin{aligned} &\|\mathbf{X}_{(\tilde{S}^k)^c}\|_F + \|(\mathbf{W}_{T_0}\mathbf{X})_{(\tilde{S}^k)^c}\|_F \\ &\leq \|((\mathbf{A}^H\mathbf{A} - \mathbf{I}_L)(\mathbf{X} - \mathbf{X}^{k-1}))_{T\setminus\tilde{S}^k}\|_F \\ &+ \|(\mathbf{W}_{T_0}(\mathbf{A}^H\mathbf{A} - \mathbf{I}_L)(\mathbf{X} - \mathbf{X}^{k-1}))_{T\setminus\tilde{S}^k}\|_F \\ &+ \|(\mathbf{W}_{T_0}(\mathbf{A}^H\mathbf{A} - \mathbf{I}_L)(\mathbf{X} - \mathbf{X}^{k-1}))_{\Delta S\setminus T}\|_F \\ &+ \|((\mathbf{A}^H\mathbf{A} - \mathbf{I}_L)(\mathbf{X} - \mathbf{X}^{k-1}))_{\Delta S\setminus T}\|_F \\ &+ \|(\mathbf{A}^H\mathbf{E})_{T\setminus\tilde{S}^k}\|_F + \|(\mathbf{W}_{T_0}\mathbf{A}^H\mathbf{E})_{T\setminus\tilde{S}^k}\|_F \\ &+ \|(\mathbf{A}^H\mathbf{E})_{\Delta S\setminus T}\|_F + \|(\mathbf{W}_{T_0}\mathbf{A}^H\mathbf{E})_{\Delta S\setminus T}\|_F \\ &\leq \sqrt{2}\|((\mathbf{A}^H\mathbf{A} - \mathbf{I}_L)(\mathbf{X} - \mathbf{X}^{k-1}))_{T\cup\Delta S}\|_F \\ &+ \sqrt{2}\|(\mathbf{W}_{T_0}(\mathbf{A}^H\mathbf{A} - \mathbf{I}_L)(\mathbf{X} - \mathbf{X}^{k-1}))_{T\cup\Delta S}\|_F \\ &+ \sqrt{2}\|(\mathbf{A}^H\mathbf{E})_{T\cup\Delta S}\|_F + \sqrt{2}\|(\mathbf{W}_{T_0}\mathbf{A}^H\mathbf{E})_{T\cup\Delta S}\|_F \\ &\leq \sqrt{2}(1+\omega)\delta_{3s}\|\mathbf{X} - \mathbf{X}^{k-1}\|_F \\ &+ (1+\omega)\sqrt{2(1+\delta_{3s})}\|\mathbf{E}\|_F. \end{aligned} \tag{A.70}$$

We can obtain the relationship between $\|\mathbf{X}_{(\tilde{S}^k)^c}\|_F$ $\|(\mathbf{W}_{T_0}\mathbf{X})_{(\tilde{S}^k)^c}\|_F$:

$$\eta\omega\sqrt{s_2}\|\mathbf{X}_{(\tilde{S}^k)^c}\|_F \leq \|(\mathbf{W}_{T_0}\mathbf{X})_{(\tilde{S}^k)^c}\|_F. \tag{A.71}$$

Combining (A.70) and (A.71), we have

$$\begin{aligned} \|\mathbf{X}_{(\tilde{S}^k)^c}\|_F &\leq \frac{\sqrt{2}(1+\omega)\delta_{3s}}{1+\eta\omega\sqrt{s_2}}\|\mathbf{X} - \tilde{\mathbf{X}}\|_F \\ &+ \frac{(1+\omega)\sqrt{2(1+\delta_{2s})}}{1+\eta\omega\sqrt{s_2}}\|\mathbf{E}\|_F. \end{aligned} \tag{A.72}$$

Noting the definition of ν_1, we complete the proof of Lemma 3.

Appendix C
Proof of Lemma 4

Proof: the RIC δ_t can be expressed as [25]

$$\delta_t = \max_{S\subseteq\{1,2,\ldots,N\},|S|\leq t} \|\mathbf{A}_S^*\mathbf{A}_S - \mathbf{I}\|_{2\to 2}, \tag{A.73}$$

where

$$\|\mathbf{A}_S^*\mathbf{A}_S - \mathbf{I}\|_{2\to 2} = \sup_{\mathbf{a}\in\mathbb{R}^{|S|}\setminus\{\mathbf{0}\}} \frac{\|(\mathbf{A}_S^*\mathbf{A}_S - \mathbf{I})\,\mathbf{a}\|_2}{\|\mathbf{a}\|_2}. \tag{A.74}$$

Let $S = \text{supp}(\mathbf{u}) \cup \text{supp}(\mathbf{v})$, then $|S| \leq t$. Let $\mathbf{u}_{|S}, \mathbf{v}_{|S}$ denote respectively the S-dimensional sub-vectors of $\mathbf{u}$ and $\mathbf{v}$ obtained by only keeping the components indexed by S. We have

$$\begin{aligned} &|\langle\mathbf{u}, \left(\mathbf{W}_{T_0} - \mathbf{W}_{T_0}\mathbf{A}^H\mathbf{A}\right)\mathbf{v}\rangle| \\ &= |\langle\mathbf{W}_{T_0}\mathbf{u}, \mathbf{v}\rangle - \langle\mathbf{A}\mathbf{W}_{T_0}\mathbf{u}, \mathbf{A}\mathbf{v}\rangle| \\ &= |\langle\mathbf{W}_{T_0}\mathbf{u}_{|S}, \left(\mathbf{I}_L - \mathbf{A}_S^H\mathbf{A}_S\right)\mathbf{v}_{|S}\rangle| \\ &\leq \|\mathbf{W}_{T_0}\mathbf{u}_{|S}\|_2 \left\|\left(\mathbf{I}_L - \mathbf{A}_S^H\mathbf{A}_S\right)\mathbf{v}_{|S}\right\|_2 \\ &\overset{(A.74)}{\leq} \|\mathbf{W}_{T_0}\mathbf{u}_{|S}\|_2\|\mathbf{I}_L - \mathbf{A}_S^H\mathbf{A}_S\|_{2\to 2}\|\mathbf{v}_{|S}\|_2 \\ &\overset{(A.73)}{\leq} \omega\delta_t\|\mathbf{u}_{|T}\|_2\|\mathbf{v}_{|S}\|_2 \end{aligned}$$

$$= \quad \omega\delta_t \|\mathbf{u}\|_2 \|\mathbf{v}\|_2, \tag{A.75}$$

moreover, we have

$$\begin{aligned} & \| \left(\left(\mathbf{W}_{T_0} - \mathbf{W}_{T_0} \mathbf{A}^H \mathbf{A} \right) \mathbf{v} \right)_U \|_2^2 \\ = \; & \langle \left(\left(\mathbf{W}_{T_0} - \mathbf{W}_{T_0} \mathbf{A}^H \mathbf{A} \right) \mathbf{v} \right)_U , \\ & \left(\mathbf{W}_{T_0} - \mathbf{W}_{T_0} \mathbf{A}^H \mathbf{A} \right) \mathbf{v} \rangle \\ \overset{(S.37)}{\leq} \; & \delta_t \left\| \left(\left(\mathbf{W}_{T_0} - \mathbf{A}^H \mathbf{A} \right) \mathbf{v} \right)_U \right\|_2 \|\mathbf{v}\|_2 \end{aligned} \tag{A.76}$$

which completes the proof of Lemma 4.

Appendix D
Proof of Lemma 5

Proof: The lemma follows trivially from the fact that

$$\begin{aligned} & \left\| \left(\mathbf{W}_{T_0} \mathbf{A}^H \mathbf{e} \right)_U \right\|_2^2 \\ & = \left\langle \mathbf{W}_{T_0} \mathbf{A}^H \mathbf{e}, \left(\mathbf{W}_{T_0} \mathbf{A}^H \mathbf{e} \right)_U \right\rangle \\ & = \left\langle \mathbf{e}, \mathbf{W}_{T_0} \mathbf{A} \left(\left(\mathbf{A}^H \mathbf{e} \right)_U \right) \right\rangle \\ & \leq \|\mathbf{e}\|_2 \left\| \mathbf{W}_{T_0} \mathbf{A} \left(\left(\mathbf{A}^H \mathbf{e} \right)_U \right) \right\|_2 \\ & \overset{(7)}{\leq} \|\mathbf{e}'\|_2 \, \omega \sqrt{1+\delta_u} \left\| \left(\mathbf{A}^H \mathbf{e} \right)_U \right\|_2 . \end{aligned} \tag{A.77}$$

Appendix E
Proof of Lemma 6

Proof: Due to the orthogonality, the residue $\mathbf{Y} - \mathbf{A}\tilde{\mathbf{X}}$ is orthogonal to the space $\mathbf{A}\mathbf{Z}$. This means that for all $\mathbf{Z} \in \mathbb{C}^{L \times N}$ with $\mathrm{supp}(\mathbf{Z}) \subseteq S$,

$$\langle \mathbf{A}(\mathbf{Y} - \mathbf{A}\tilde{\mathbf{X}}), \mathbf{Z} \rangle = \mathbf{0}. \tag{A.78}$$

Let $\tilde{\mathbf{X}}'$ be the solution of the least squares problem $\arg\min_{\mathbf{Z}} \{ \|\mathbf{Y}' - \mathbf{A}\mathbf{Z}\|_F , \mathrm{supp}(\mathbf{Z}) \subseteq S \}$, where $\mathbf{Y}' = \frac{\mathbf{A}\mathbf{W}_{T_0}\mathbf{X}_{T_0}}{\omega} + \mathbf{E}$. We have

$$\tilde{\mathbf{X}}' = \frac{\mathbf{W}_{T_0}\tilde{\mathbf{X}}}{\omega}. \tag{A.79}$$

Then, by (A.78), we have

$$\begin{aligned} \mathbf{0} \; & = \; \left\langle \frac{\mathbf{A}\mathbf{W}_{T_0}\mathbf{X}_{T_0}}{\omega} + \mathbf{E} - \mathbf{A}\frac{\mathbf{W}_{T_0}\tilde{X}}{\omega}, \mathbf{A}\mathbf{Z} \right\rangle \\ & = \; \left\langle \mathbf{W}_{T_0}\mathbf{X} - \mathbf{W}_{T_0}\tilde{\mathbf{X}}, \mathbf{A}^H\mathbf{A}\mathbf{Z} \right\rangle + \omega \left\langle \mathbf{E}, \mathbf{A}\mathbf{Z} \right\rangle . \end{aligned} \tag{A.80}$$