Supplementary Materials for Compressed Smooth Sparse Decomposition*

Appendices

Appendix A. Proof of Theorem 2.1.

We prove Theorem 2.1 by contradiction: Assume there exists two different decompositions for the same smooth plus sparse signal $\boldsymbol{y}$, i.e., $\boldsymbol{y} = \boldsymbol{m}_1 + \boldsymbol{a}_1 = \boldsymbol{m}_2 + \boldsymbol{a}_2$, where $\boldsymbol{m}_1 \neq \boldsymbol{m}_2$ and $\boldsymbol{a}_1 \neq \boldsymbol{a}_2$. Then,

$$\boldsymbol{m}_1 - \boldsymbol{m}_2 = -\boldsymbol{a}_1 + \boldsymbol{a}_2 \tag{3}$$

Since $\boldsymbol{m}_1 \neq \boldsymbol{m}_2$, we can normalize both side by $\|\boldsymbol{m}_1 - \boldsymbol{m}_2\|_2$ and denote $\widetilde{\boldsymbol{m}} = (\boldsymbol{m}_1 - \boldsymbol{m}_2)/\|\boldsymbol{m}_1 - \boldsymbol{m}_2\|_2$ and $\widetilde{\boldsymbol{a}} = (-\boldsymbol{a}_1 + \boldsymbol{a}_2)/\|\boldsymbol{m}_1 - \boldsymbol{m}_2\|_2$. Notice that $\widetilde{\boldsymbol{m}}$ is in the column space of $\boldsymbol{B}$ which is spanned by columns of the $\boldsymbol{U}$ (recall that $\boldsymbol{B} = \boldsymbol{U\Sigma V}^{\mathrm{T}}$), i.e., $\widetilde{\boldsymbol{m}} = \boldsymbol{Ux}$, where $\boldsymbol{x}$ is the coefficient vector, $\boldsymbol{x} \in \mathbb{R}^r$. Since $\|\widetilde{\boldsymbol{m}}\|_2 = 1$, we have $\|\boldsymbol{x}\|_2 = 1$. We can bound each element in $\widetilde{\boldsymbol{m}}$ as follows:

$$|\widetilde{m}_i| = \boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{Ux} \leq \left\|\boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{U}\right\|_2 \|\boldsymbol{x}\|_2 \leq \max_{i\in\{1,..,r\}} \left\|\boldsymbol{U}^{\mathrm{T}}\boldsymbol{e}_i\right\|_2 \leq \sqrt{\frac{\mu(\boldsymbol{B})r}{n}} < \sqrt{\frac{1}{2ls}}, \qquad \forall i \in \{1,\dots,n\}.$$

According to Definition 2.1, $\|\boldsymbol{a}_1\|_0 \leq ls$ and $\|\boldsymbol{a}_2\|_0 \leq ls$. Therefore, $\|\widetilde{\boldsymbol{a}}\|_0 \leq 2ls$. Moreover, according to Eq. (3), we conclude that $\|\widetilde{\boldsymbol{m}}\|_0 \leq 2ls$. Denote the support of $\widetilde{\boldsymbol{m}}$ as $\mathbf{T}_{\widetilde{\boldsymbol{m}}}$, we have,

$\|\widetilde{\boldsymbol{m}}\|_2 = \sqrt{\sum_{i\in\mathbf{T}_{\widetilde{\boldsymbol{m}}}}|\widetilde{m}_i|^2} < 1$ which is a contradiction. ■

Appendix B. Proof of Theorem 2.2.

We prove the Theorem 2.2 with contradiction: Assume there exists another vector $\boldsymbol{y} = \boldsymbol{B\theta} + \boldsymbol{B}_a\boldsymbol{\theta}_a \in MS_{r,s,\mu,l}$ such that $\boldsymbol{Ay} = \boldsymbol{b}$ and $\boldsymbol{y} \neq \boldsymbol{y}_0$. Then, $\boldsymbol{z} = \boldsymbol{y} - \boldsymbol{y}_0 = \boldsymbol{B}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + \boldsymbol{B}_a(\boldsymbol{\theta}_a - \boldsymbol{\theta}_{a0})$ is a nonzero vector. Since $\|\boldsymbol{\theta}_a - \boldsymbol{\theta}_{a0}\|_1 \leq \|\boldsymbol{\theta}_a\|_1 + \|\boldsymbol{\theta}_{a0}\|_1 \leq 2s$, by Definition 2.3, we have that $\boldsymbol{z} \in MS_{r,2s,\mu,l}$. Therefore, $0 = \|\boldsymbol{Az}\|_2 \geq (1 - \delta_{r,2s,\mu,l})\|\boldsymbol{z}\|_2 > 0$, which is a contradiction. Notice that the proof is inspired by the proof of Lemma 3.1 in Candes and Tao (2005). ■

Appendix C. Proof of Theorem 2.3.

This proof is inspired by Baraniuk et al. (2008) and Tanner and Vary (2020).

We will first derive the RIC for a fixed subspace $MS_{r,\mathbf{T},\mu,l}$of $MS_{r,s,\mu,l}$ when $\boldsymbol{\theta}_a$ is restricted in a fixed subspace $\mathbf{T}$ with the fixed support such that the number of non-zero elements is s.

$MS_{r,\mathbf{T},\mu,l} = \{\boldsymbol{y} \in \mathbb{R}^n \mid \boldsymbol{y} = \boldsymbol{B\theta} + \boldsymbol{B}_a\boldsymbol{\theta}_a, \boldsymbol{B}_a \in \mathbb{R}^{n\times q}, l(\boldsymbol{B}_a) = l, \boldsymbol{\theta}_a \in \mathbf{T}, \boldsymbol{B} \in \mathbb{R}^{n\times r}, \mu(\boldsymbol{B}) = \mu < n(2rsl)^{-1}, \boldsymbol{\theta} \in \mathbb{R}^r\}$.

Then, we use a covering argument that counts over all possible sparse subspaces $\mathbf{T}$ with support less than or equal to s. Finally, we can derive the RIC for $MS_{r,s,\mu,l}$.

The following lemma describes the RIC for a fixed subspace $MS_{r,\mathbf{T},\mu,l}$ and is proved in Appendix E.

* No data ethics considerations are foreseen related to this paper.

Lemma 6.1. *RIC for a fixed subspace $MS_{r,\boldsymbol{T},\mu,l}$. Let $\boldsymbol{A} \in \mathbb{R}^{p\times n}$ be a matrix from the families described in Definition 3.5. Further, assume $\mu < n(2rsl)^{-1}$ and basis matrix $\boldsymbol{B}_a$ for sparse signal component satisfies the RIP with RIC $\delta_{\boldsymbol{B}_a,s} \in (0,1)$, i.e., $\delta_{\boldsymbol{B}_a,s}$ is the smallest positive constant such that*

$$\left(1-\delta_{\boldsymbol{B}_a,s}\right)\|\boldsymbol{\theta}_a\|_2 \le \|\boldsymbol{B}_a\boldsymbol{\theta}_a\|_2 \le \left(1+\delta_{\boldsymbol{B}_a,s}\right)\|\boldsymbol{\theta}_a\|_2, \qquad \forall\, \boldsymbol{\theta}_a \in \{\,\boldsymbol{\theta}_a \in \mathbb{R}^q \,|\, \|\boldsymbol{\theta}_a\|_0 \le s\}.$$

For a given $\delta \in (0,1)$, there exists a constant $c_0 > 0$ depending only on δ, such that the RIC for $MS_{r,\boldsymbol{T},\mu,l}$ is upper bounded by δ with the probability of at least $1-2\left(\frac{24}{\delta}\tau_1\right)^r\left(\frac{24}{\delta}\tau_0\right)^s e^{-pc_0(\delta/2)}$, where $\eta=\sqrt{\frac{\mu rsl}{n}}$, $\tau_0=\frac{1}{\sqrt{(1-\delta_{\boldsymbol{B}_a,s})(1-\eta^2)}}$, $\tau_1=\left\|\boldsymbol{B}^\dagger\right\|_2\left(1+\frac{1}{\sqrt{1-\eta^2}}\right)$, and $\boldsymbol{B}^\dagger=\left(\boldsymbol{B}^T\boldsymbol{B}\right)^{-1}\boldsymbol{B}^T$. ■

Notice that for a fixed subspace $MS_{r,\mathbf{T},\mu,l}$, the RIP will fail with probability less than or equal to $2\left(\frac{24}{\delta}\tau_1\right)^r\left(\frac{24}{\delta}\tau_0\right)^s e^{-pc_0(\delta/2)}$. Since there are $\binom{n}{s} \le \left(\frac{en}{s}\right)^s$ such subspaces, the probability to fail for $MS_{r,s,\mu,l}$, which is a combination of those $\binom{n}{s}$ subspaces, will be less than or equal to

$$\binom{n}{s}2\left(\frac{24}{\delta}\tau_1\right)^r\left(\frac{24}{\delta}\tau_0\right)^s e^{-pc_0\left(\frac{\delta}{2}\right)} \le \exp\left(-c_0\left(\frac{\delta}{2}\right)p+\ln 2+r\ln\frac{24}{\delta}\tau_1+s\left(1+\ln\frac{24}{\delta}\tau_0+\ln\frac{n}{s}\right)\right).$$

Then, for any give δ, there exist $c_1, c_2 > 0$, such that the probability to fail for $MS_{r,s,\mu,l}$ is less than or equal to $\exp(-c_1 p)$, provided that $p \ge c_2\left(\ln 2+r\ln\frac{24}{\delta}\tau_1+s\left(1+\ln\frac{24}{\delta}\tau_0+\ln\frac{n}{s}\right)\right)$, where $c_2=\left[c_0\left(\frac{\delta}{2}\right)-c_1\right]^{-1}$. This finishes the proof. ■

Appendix D. Proof of Theorem 2.4.

The proof of Theorem 2.4 is inspired by Candes et al. (2006) and Tanner and Vary (2020). Assume that in Problem (1), ϵ_1 is properly chosen such that Problem (1) is feasible. In the following discussion, we will use $(\cdot)^*$ to denote the optimal solution of Problem (1) and $(\cdot)_0$ to denote the signal we wish to recover. Let $\boldsymbol{R}=\boldsymbol{X}^*-\boldsymbol{X}_0=\boldsymbol{R}^{\boldsymbol{m}}+\boldsymbol{R}^{\boldsymbol{a}}$ where $\boldsymbol{R}^{\boldsymbol{m}}=\boldsymbol{m}-\boldsymbol{m}_0=\boldsymbol{B}\boldsymbol{\theta}^*-\boldsymbol{B}\boldsymbol{\theta}_0$ and $\boldsymbol{R}^{\boldsymbol{a}}=\boldsymbol{B}_a\boldsymbol{\theta}_a^*-\boldsymbol{B}_a\boldsymbol{\theta}_{a0}$ are the residual of the smooth and sparse signal component, respectively.

Let $\boldsymbol{h}=\boldsymbol{\theta}_a^*-\boldsymbol{\theta}_{a0}=\boldsymbol{h}_{\mathbf{T}_0}+\boldsymbol{h}_{\mathbf{T}_0^c}$, where $\mathbf{T}_0$ is the support of $\boldsymbol{\theta}_{a0}$ and $\boldsymbol{h}_{\mathbf{T}_0}$ denotes the projection of $\boldsymbol{h}$ onto $\mathbf{T}_0$ such that

$$\boldsymbol{h}_{\mathbf{T}_0}(t)=\begin{cases} t, & \text{if } t\in\mathbf{T}_0 \\ 0, & \text{otherwise}\end{cases},$$

and $\mathbf{T}_0^c$ denotes the complementary set of $\mathbf{T}_0$.

Since $\boldsymbol{\theta}_{a0}$ is feasible and $\boldsymbol{\theta}_a^*$ is the optimal solution of Problem (1), we must have $\|\boldsymbol{\theta}_a^*\|_1 \le \|\boldsymbol{\theta}_{a0}\|_1$, which is equivalent to $\left\|\boldsymbol{\theta}_{a0}+\boldsymbol{h}_{\mathbf{T}_0}+\boldsymbol{h}_{\mathbf{T}_0^c}\right\|_1 \le \|\boldsymbol{\theta}_{a0}\|_1$. Since $\mathbf{T}_0$ and $\mathbf{T}_0^c$ are complementary to each other, we have $\left\|\boldsymbol{\theta}_{a0}+\boldsymbol{h}_{\mathbf{T}_0}\right\|_1+\left\|\boldsymbol{h}_{\mathbf{T}_0^c}\right\|_1 \le \|\boldsymbol{\theta}_{a0}\|_1$. Since $\left\|\boldsymbol{\theta}_{a0}+\boldsymbol{h}_{\mathbf{T}_0}\right\|_1 \ge \|\boldsymbol{\theta}_{a0}\|_1-\left\|\boldsymbol{h}_{\mathbf{T}_0}\right\|_1$, we have $\|\boldsymbol{\theta}_{a0}\|_1-\left\|\boldsymbol{h}_{\mathbf{T}_0}\right\|_1+\left\|\boldsymbol{h}_{\mathbf{T}_0^c}\right\|_1 \le \|\boldsymbol{\theta}_{a0}\|_1$. Hence $\left\|\boldsymbol{h}_{\mathbf{T}_0^c}\right\|_1 \le \left\|\boldsymbol{h}_{\mathbf{T}_0}\right\|_1$. Since $\left\|\boldsymbol{h}_{\mathbf{T}_0}\right\|_1 \le \sqrt{s}\left\|\boldsymbol{h}_{\mathbf{T}_0}\right\|_2$, we have

$$\left\|\boldsymbol{h}_{\mathbf{T}_0^c}\right\|_1 \le \sqrt{s}\left\|\boldsymbol{h}_{\mathbf{T}_0}\right\|_2 \tag{4}$$

Similar to Candes et al. (2006), we order the elements of $\mathbf{T}_0^c$ in decreasing order of their magnitude and enumerate $\mathbf{T}_0^c$ as $v_1, \dots, v_{n-|\mathbf{T}_0|}$. Then, $\mathbf{T}_0^c$ is divided into subsets $\mathbf{T}_i^c$ of size M, where

$$\mathbf{T}_i^c = \{v_j : (i-1)M \le j \le iM\}.$$

Let $\boldsymbol{h}_{\mathbf{T}_i^c}$ be the projection of $\boldsymbol{h}$ onto $\mathbf{T}_i^c$, we have

$$\left\|\boldsymbol{h}_{\mathbf{T}_i^c}\right\|_0 \le M, \qquad \forall i \ge 1$$

$$\mathbf{T}_i^c \cap \mathbf{T}_j^c = \emptyset, \qquad \forall i \ne j$$

$$\left\|\boldsymbol{h}_{\mathbf{T}_{i+1}^c}\right\|_2 \le \frac{1}{\sqrt{M}}\left\|\boldsymbol{h}_{\mathbf{T}_i^c}\right\|_1, \qquad \forall i \ge 1 \tag{5}$$

where the last inequality comes from the fact that $\mathbf{T}_0^c$ is in decreasing order, such that

$$\left|\boldsymbol{h}_{\mathbf{T}_{i+1}^c}\right|_{(v)} \le \frac{1}{M}\sum_{j\in\mathbf{T}_i^c}\left|\boldsymbol{h}_{\mathbf{T}_i^c}\right|_{(j)}. \ \forall v \in \mathbf{T}_{i+1}^c$$

Define $\boldsymbol{R}^{\boldsymbol{a}}_{\mathbf{T}_i^c} = \boldsymbol{B}_a \boldsymbol{h}_{\mathbf{T}_i^c}$, $\boldsymbol{R}^{\boldsymbol{a}}_{\mathbf{T}_0} = \boldsymbol{B}_a \boldsymbol{h}_{\mathbf{T}_0}$ and combine Eq. (4) and Eq. (5) , we have

$$\sum_{j\ge2}\left\|\boldsymbol{R}^{\boldsymbol{a}}_{\mathbf{T}_j^c}\right\|_2 \le \sum_{j\ge2}\sqrt{1+\delta_{\boldsymbol{B}_a,M}}\left\|\boldsymbol{h}_{\mathbf{T}_j^c}\right\|_2 \le_{(a)} \sum_{j\ge1}\frac{\sqrt{1+\delta_{\boldsymbol{B}_a,M}}\left\|\boldsymbol{h}_{\mathbf{T}_j^c}\right\|_1}{\sqrt{M}} = \frac{\sqrt{1+\delta_{\boldsymbol{B}_a,M}}\left\|\boldsymbol{h}_{\mathbf{T}_0^c}\right\|_1}{\sqrt{M}}$$
$$\le_{(b)} \frac{\sqrt{s}\sqrt{1+\delta_{\boldsymbol{B}_a,M}}\left\|\boldsymbol{h}_{\mathbf{T}_0}\right\|_2}{\sqrt{M}} \le \frac{\sqrt{s}\sqrt{1+\delta_{\boldsymbol{B}_a,M}}\left\|\boldsymbol{R}^{\boldsymbol{a}}_{\mathbf{T}_0}\right\|_2}{\sqrt{M}\sqrt{1-\delta_{\boldsymbol{B}_a,M}}},$$

where (a) follows Eq.(5), (b) follows Eq. (4) and the RIP property of $\boldsymbol{B}_a$ is used since $\left\|\boldsymbol{h}_{\mathbf{T}_j^c}\right\|_0 \le M$.

Denote $\gamma = \sqrt{\frac{1+\delta_{\boldsymbol{B}_a,M+s}}{1-\delta_{\boldsymbol{B}_a,M+s}}}$ (A tighter bound can be achieved by using $\sqrt{\frac{1+\delta_{\boldsymbol{B}_a,M}}{1-\delta_{\boldsymbol{B}_a,M}}}$. However, we adopt $\delta_{\boldsymbol{B}_a,M+s}$ instead of $\delta_{\boldsymbol{B}_a,M}$ for simplicity in the following proof), we have

$$\sum_{j\ge2}\left\|\boldsymbol{R}^{\boldsymbol{a}}_{\mathbf{T}_j^c}\right\|_2 \le \sqrt{\frac{s}{M}}\gamma\left\|\boldsymbol{R}^{\boldsymbol{a}}_{\mathbf{T}_0}\right\|_2 \tag{6}$$

Next, we derive the bound for $\boldsymbol{R}^{\boldsymbol{m}}$ and $\boldsymbol{R}^{\boldsymbol{a}}$ respectively.

Bound for $\boldsymbol{R}^{\boldsymbol{m}}$:

$$\begin{aligned}
&\|\boldsymbol{A}\boldsymbol{R}^{\boldsymbol{m}}\|_2^2 = |\langle \boldsymbol{A}\boldsymbol{R}^{\boldsymbol{m}}, \boldsymbol{A}(\boldsymbol{R}-\boldsymbol{R}^{\boldsymbol{a}})\rangle| \\
&= |\langle \boldsymbol{A}\boldsymbol{R}^{\boldsymbol{m}}, \boldsymbol{A}(\boldsymbol{R}-\boldsymbol{R}^{\boldsymbol{a}})\rangle| \\
&= |\langle \boldsymbol{A}\boldsymbol{R}^{\boldsymbol{m}}, \boldsymbol{A}\boldsymbol{R}\rangle + \langle \boldsymbol{A}\boldsymbol{R}^{\boldsymbol{m}}, -\boldsymbol{A}\boldsymbol{R}^{\boldsymbol{a}}\rangle| \\
&\le |\langle \boldsymbol{A}\boldsymbol{R}^{\boldsymbol{m}}, \boldsymbol{A}\boldsymbol{R}\rangle| + |\langle \boldsymbol{A}\boldsymbol{R}^{\boldsymbol{m}}, -\boldsymbol{A}\boldsymbol{R}^{\boldsymbol{a}}\rangle| \\
&= |\langle \boldsymbol{A}\boldsymbol{R}^{\boldsymbol{m}}, \boldsymbol{A}\boldsymbol{R}\rangle| + \left|\langle \boldsymbol{A}\boldsymbol{R}^{\boldsymbol{m}}, -\boldsymbol{A}\left(\boldsymbol{R}^{\boldsymbol{a}}_{\mathbf{T}_0} + \boldsymbol{R}^{\boldsymbol{a}}_{\mathbf{T}_1^c} + \textstyle\sum_{j\ge2}\boldsymbol{R}^{\boldsymbol{a}}_{\mathbf{T}_j^c}\right)\rangle\right| \\
&\le |\langle \boldsymbol{A}\boldsymbol{R}^{\boldsymbol{m}}, \boldsymbol{A}\boldsymbol{R}\rangle| + \left|\langle \boldsymbol{A}\boldsymbol{R}^{\boldsymbol{m}}, \boldsymbol{A}\left(\boldsymbol{R}^{\boldsymbol{a}}_{\mathbf{T}_0} + \boldsymbol{R}^{\boldsymbol{a}}_{\mathbf{T}_1^c}\right)\rangle\right| + \sum_{j\ge2}\left|\langle \boldsymbol{A}\boldsymbol{R}^{\boldsymbol{m}}, \boldsymbol{A}\boldsymbol{R}^{\boldsymbol{a}}_{\mathbf{T}_j^c}\rangle\right|.
\end{aligned} \tag{7}$$

In the following discussion, we will bound those terms respectively. According to Cauchy-Schwarz inequality, the first term can be bounded as

$$|\langle \boldsymbol{A}\boldsymbol{R}^{\boldsymbol{m}}, \boldsymbol{A}\boldsymbol{R}\rangle| \leq \|\boldsymbol{A}\boldsymbol{R}^{\boldsymbol{m}}\|_2 \|\boldsymbol{A}\boldsymbol{R}\|_2 \leq \sqrt{1+\delta_{r,s,\mu,l}}\epsilon_1 \|\boldsymbol{R}^{\boldsymbol{m}}\|_2, \tag{8}$$

where the last inequality comes from the RIP property and the first constraint in Problem (1).

The third term can be bounded as follows: denote $\boldsymbol{z}_1 = \boldsymbol{R}^{\boldsymbol{m}}/\|\boldsymbol{R}^{\boldsymbol{m}}\|_2$ and $\boldsymbol{z}_2 = \boldsymbol{R}^{\boldsymbol{a}}_{\mathrm{T}_j^c}/\left\|\boldsymbol{R}^{\boldsymbol{a}}_{\mathrm{T}_j^c}\right\|_2$, we have

$$\frac{\left|\langle \boldsymbol{A}\boldsymbol{R}^{\boldsymbol{m}}, \boldsymbol{A}\boldsymbol{R}^{\boldsymbol{a}}_{\mathrm{T}_j^c}\rangle\right|}{\|\boldsymbol{R}^{\boldsymbol{m}}\|_2 \left\|\boldsymbol{R}^{\boldsymbol{a}}_{\mathrm{T}_j^c}\right\|_2} = |\langle \boldsymbol{A}\boldsymbol{z}_1, \boldsymbol{A}\boldsymbol{z}_2\rangle| = \frac{1}{4}\left|\|\boldsymbol{A}(\boldsymbol{z}_1+\boldsymbol{z}_2)\|_2^2 - \|\boldsymbol{A}(\boldsymbol{z}_1-\boldsymbol{z}_2)\|_2^2\right|$$

$$\leq_{(a)} \frac{1}{4}\max\begin{Bmatrix} \left|\left(1+\delta_{r,M,\mu,l}\right)\|\boldsymbol{z}_1+\boldsymbol{z}_2\|_2^2 - \left(1-\delta_{r,M,\mu,l}\right)\|\boldsymbol{z}_1-\boldsymbol{z}_2\|_2^2\right|, \\ \left|\left(1+\delta_{r,M,\mu,l}\right)\|\boldsymbol{z}_1-\boldsymbol{z}_2\|_2^2 - \left(1-\delta_{r,M,\mu,l}\right)\|\boldsymbol{z}_1+\boldsymbol{z}_2\|_2^2\right| \end{Bmatrix}$$

$$= \left|\delta_{r,M,\mu,l} + \langle \boldsymbol{z}_1, \boldsymbol{z}_2\rangle\right| \leq_{(b)} \delta_{r,M,\mu,l} + \frac{\eta_1}{1-\eta_1^2},$$

where $\eta_1 = \sqrt{\frac{\mu r M l}{n}}$; inequality (a) follows the RIP property since $\boldsymbol{z}_1 + \boldsymbol{z}_2 \in MS_{r,M,\mu,l}$ and $\boldsymbol{z}_1 - \boldsymbol{z}_2 \in MS_{r,M,\mu,l}$; inequality (b) follows Eq. (26) in the proof of Lemma 6.2 and $\|\boldsymbol{z}_1\|_2 = \|\boldsymbol{z}_2\|_2 = 1$, $\langle \boldsymbol{z}_1, \boldsymbol{z}_2\rangle \leq \frac{\eta_1}{1-\eta_1^2}\|\boldsymbol{z}_1\|_2\|\boldsymbol{z}_2\|_2 = \frac{\eta_1}{1-\eta_1^2}$, provided that $M \leq 2s$.

Therefore,

$$\left|\langle \boldsymbol{A}\boldsymbol{R}^{\boldsymbol{m}}, \boldsymbol{A}\boldsymbol{R}^{\boldsymbol{a}}_{\mathrm{T}_j^c}\rangle\right| \leq \left(\delta_{r,M,\mu,l} + \frac{\eta_1}{1-\eta_1^2}\right)\|\boldsymbol{R}^{\boldsymbol{m}}\|_2 \left\|\boldsymbol{R}^{\boldsymbol{a}}_{\mathrm{T}_j^c}\right\|_2. \tag{9}$$

Similarly, the second term can be bounded as

$$\left|\langle \boldsymbol{A}\boldsymbol{R}^{\boldsymbol{m}}, \boldsymbol{A}\left(\boldsymbol{R}^{\boldsymbol{a}}_{\mathrm{T}_0} + \boldsymbol{R}^{\boldsymbol{a}}_{\mathrm{T}_1^c}\right)\rangle\right| \leq \left(\delta_{r,M+s,\mu,l} + \frac{\eta_2}{1-\eta_2^2}\right)\|\boldsymbol{R}^{\boldsymbol{m}}\|_2 \left\|\boldsymbol{R}^{\boldsymbol{a}}_{\mathrm{T}_0} + \boldsymbol{R}^{\boldsymbol{a}}_{\mathrm{T}_1^c}\right\|_2, \tag{10}$$

where $\eta_2 = \sqrt{\frac{\mu r(M+s)l}{n}}$, provided that $M \leq s$.

Plugging Eq. (8), Eq. (9) and Eq. (10) into Eq. (7), we have

$$\|\boldsymbol{A}\boldsymbol{R}^{\boldsymbol{m}}\|_2^2 \leq \|\boldsymbol{R}^{\boldsymbol{m}}\|_2\left(\sqrt{1+\delta_{r,s,\mu,l}}\epsilon_1 + \left(\delta_{r,M+s,\mu,l} + \frac{\eta_2}{1-\eta_2^2}\right)\left\|\boldsymbol{R}^{\boldsymbol{a}}_{\mathrm{T}_0} + \boldsymbol{R}^{\boldsymbol{a}}_{\mathrm{T}_1^c}\right\|_2 + \left(\delta_{r,M,\mu,l} + \frac{\eta_1}{1-\eta_1^2}\right)\left\|\boldsymbol{R}^{\boldsymbol{a}}_{\mathrm{T}_j^c}\right\|_2\right)$$

$$\leq_{(a)} \|\boldsymbol{R}^{\boldsymbol{m}}\|_2\left(\sqrt{1+\delta_{r,s,\mu,l}}\epsilon_1 + \left(\delta_{r,M+s,\mu,l} + \frac{\eta_2}{1-\eta_2^2}\right)\left\|\boldsymbol{R}^{\boldsymbol{a}}_{\mathrm{T}_0} + \boldsymbol{R}^{\boldsymbol{a}}_{\mathrm{T}_1^c}\right\|_2 + \left(\delta_{r,M,\mu,l} + \frac{\eta_1}{1-\eta_1^2}\right)\sqrt{\frac{s}{M}}\gamma\left\|\boldsymbol{R}^{\boldsymbol{a}}_{\mathrm{T}_0}\right\|_2\right),$$

where inequality (a) follows Eq. (6).

According to the RIP property, we have

$$\left(1-\delta_{r,s,\mu,l}\right)\|\boldsymbol{R}^{\boldsymbol{m}}\|_2^2$$

$$\leq \|\boldsymbol{R}^{\boldsymbol{m}}\|_2\left(\sqrt{1+\delta_{r,s,\mu,l}}\epsilon_1 + \left(\delta_{r,M+s,\mu,l} + \frac{\eta_2}{1-\eta_2^2}\right)\left\|\boldsymbol{R}^{\boldsymbol{a}}_{\mathrm{T}_0} + \boldsymbol{R}^{\boldsymbol{a}}_{\mathrm{T}_1^c}\right\|_2 + \left(\delta_{r,M,\mu,l} + \frac{\eta_1}{1-\eta_1^2}\right)\sqrt{\frac{s}{M}}\gamma\left\|\boldsymbol{R}^{\boldsymbol{a}}_{\mathrm{T}_0}\right\|_2\right).$$

Consequently, we have

$$\|\boldsymbol{R}^{\boldsymbol{m}}\|_2 \leq \frac{\sqrt{1+\delta_{r,s,\mu,l}}\epsilon_1 + \left(\left(\delta_{r,M,\mu,l} + \frac{\eta_1}{1-\eta_1^2}\right)\sqrt{\frac{s}{M}}\gamma^2 + \left(\delta_{r,M+s,\mu,l} + \frac{\eta_2}{1-\eta_2^2}\right)\right)\left\|\boldsymbol{R}_{\mathbf{T}_0}^{\boldsymbol{a}} + \boldsymbol{R}_{\mathbf{T}_1^c}^{\boldsymbol{a}}\right\|_2}{\left(1-\delta_{r,s,\mu,l}\right)}. \tag{11}$$

where the inequality follows from:

$$\begin{aligned}\left\|\boldsymbol{R}_{\mathbf{T}_0}^{\boldsymbol{a}}\right\|_2 &= \left\|\boldsymbol{B}_a \boldsymbol{h}_{\mathbf{T}_0}\right\|_2 \leq_{(a)} \sqrt{1+\delta_{\boldsymbol{B}_a,s}}\left\|\boldsymbol{h}_{\mathbf{T}_0}\right\|_2 \leq_{(b)} \sqrt{1+\delta_{\boldsymbol{B}_a,s}}\left\|\boldsymbol{h}_{\mathbf{T}_0} + \boldsymbol{h}_{\mathbf{T}_1^c}\right\|_2 \\ &\leq_{(c)} \sqrt{\frac{1+\delta_{\boldsymbol{B}_a,s}}{1-\delta_{\boldsymbol{B}_a,M+s}}}\left\|\boldsymbol{B}_a\left(\boldsymbol{h}_{\mathbf{T}_0} + \boldsymbol{h}_{\mathbf{T}_1^c}\right)\right\|_2 \leq \gamma\left\|\boldsymbol{R}_{\mathbf{T}_0}^{\boldsymbol{a}} + \boldsymbol{R}_{\mathbf{T}_1^c}^{\boldsymbol{a}}\right\|_2,\end{aligned} \tag{12}$$

where inequalities (a) and (c) follow the RIP property of $\boldsymbol{B}_a$ and inequality (b) follows that $\mathbf{T}_0 \cap \mathbf{T}_1^c = \emptyset$.

Bound for $\boldsymbol{R}^{\boldsymbol{a}}$:

$$\begin{aligned}&\left\|\boldsymbol{A}\left(\boldsymbol{R}_{\mathbf{T}_0}^{\boldsymbol{a}} + \boldsymbol{R}_{\mathbf{T}_1^c}^{\boldsymbol{a}}\right)\right\|_2^2 = \left|\left\langle \boldsymbol{A}\left(\boldsymbol{R}_{\mathbf{T}_0}^{\boldsymbol{a}} + \boldsymbol{R}_{\mathbf{T}_1^c}^{\boldsymbol{a}}\right), \boldsymbol{A}\left(\boldsymbol{R}_{\mathbf{T}_0}^{\boldsymbol{a}} + \boldsymbol{R}_{\mathbf{T}_1^c}^{\boldsymbol{a}} - \boldsymbol{R} + \boldsymbol{R}\right)\right\rangle\right| \\ &= \left|\left\langle \boldsymbol{A}\left(\boldsymbol{R}_{\mathbf{T}_0}^{\boldsymbol{a}} + \boldsymbol{R}_{\mathbf{T}_1^c}^{\boldsymbol{a}}\right), \boldsymbol{A}\boldsymbol{R}\right\rangle\right| + \left|\left\langle \boldsymbol{A}\left(\boldsymbol{R}_{\mathbf{T}_0}^{\boldsymbol{a}} + \boldsymbol{R}_{\mathbf{T}_1^c}^{\boldsymbol{a}}\right), -\boldsymbol{A}\left(\boldsymbol{R}^{\boldsymbol{m}} + \textstyle\sum_{j\geq 2}\boldsymbol{R}_{\mathbf{T}_j^c}^{\boldsymbol{a}}\right)\right\rangle\right| \\ &\leq \left|\left\langle \boldsymbol{A}\left(\boldsymbol{R}_{\mathbf{T}_0}^{\boldsymbol{a}} + \boldsymbol{R}_{\mathbf{T}_1^c}^{\boldsymbol{a}}\right), \boldsymbol{A}\boldsymbol{R}\right\rangle\right| + \left|\left\langle \boldsymbol{A}\boldsymbol{R}^{\boldsymbol{m}}, \boldsymbol{A}\left(\boldsymbol{R}_{\mathbf{T}_0}^{\boldsymbol{a}} + \boldsymbol{R}_{\mathbf{T}_1^c}^{\boldsymbol{a}}\right)\right\rangle\right| + \textstyle\sum_{j\geq 2}\left|\left\langle \boldsymbol{A}\left(\boldsymbol{R}_{\mathbf{T}_0}^{\boldsymbol{a}} + \boldsymbol{R}_{\mathbf{T}_1^c}^{\boldsymbol{a}}\right), \boldsymbol{A}\boldsymbol{R}_{\mathbf{T}_j^c}^{\boldsymbol{a}}\right\rangle\right|\end{aligned} \tag{13}$$

In the following discussion, we will bound those terms respectively. According to Cauchy-Schwarz inequality, the first term can be bounded as

$$\left|\left\langle \boldsymbol{A}\left(\boldsymbol{R}_{\mathbf{T}_0}^{\boldsymbol{a}} + \boldsymbol{R}_{\mathbf{T}_1^c}^{\boldsymbol{a}}\right), \boldsymbol{A}\boldsymbol{R}\right\rangle\right| \leq \left\|\boldsymbol{A}\left(\boldsymbol{R}_{\mathbf{T}_0}^{\boldsymbol{a}} + \boldsymbol{R}_{\mathbf{T}_1^c}^{\boldsymbol{a}}\right)\right\|_2 \|\boldsymbol{A}\boldsymbol{R}\|_2 \leq \sqrt{1+\delta_{r,M+s,\mu,l}}\epsilon_1 \left\|\boldsymbol{R}_{\mathbf{T}_0}^{\boldsymbol{a}} + \boldsymbol{R}_{\mathbf{T}_1^c}^{\boldsymbol{a}}\right\|_2, \tag{14}$$

where the last inequality comes from the RIP property and the first constraint in Problem (1).

The third term can be bounded as follows: denote $\boldsymbol{z}_2 = \boldsymbol{R}_{\mathbf{T}_j^c}^{\boldsymbol{a}} / \left\|\boldsymbol{R}_{\mathbf{T}_j^c}^{\boldsymbol{a}}\right\|_2, j \geq 2$, and $\boldsymbol{z}_3 = \left(\boldsymbol{R}_{\mathbf{T}_0}^{\boldsymbol{a}} + \boldsymbol{R}_{\mathbf{T}_1^c}^{\boldsymbol{a}}\right) / \left\|\boldsymbol{R}_{\mathbf{T}_0}^{\boldsymbol{a}} + \boldsymbol{R}_{\mathbf{T}_1^c}^{\boldsymbol{a}}\right\|_2$, we have

$$\begin{aligned}&\frac{\left|\left\langle \boldsymbol{A}\left(\boldsymbol{R}_{\mathbf{T}_0}^{\boldsymbol{a}} + \boldsymbol{R}_{\mathbf{T}_1^c}^{\boldsymbol{a}}\right), \boldsymbol{A}\boldsymbol{R}_{\mathbf{T}_j^c}^{\boldsymbol{a}}\right\rangle\right|}{\left\|\boldsymbol{R}_{\mathbf{T}_0}^{\boldsymbol{a}} + \boldsymbol{R}_{\mathbf{T}_1^c}^{\boldsymbol{a}}\right\|_2 \left\|\boldsymbol{R}_{\mathbf{T}_j^c}^{\boldsymbol{a}}\right\|_2} = |\langle \boldsymbol{A}\boldsymbol{z}_3, \boldsymbol{A}\boldsymbol{z}_2\rangle| \\ &= \frac{1}{4}\left|\|\boldsymbol{A}(\boldsymbol{z}_3 + \boldsymbol{z}_2)\|_2^2 - \|\boldsymbol{A}(\boldsymbol{z}_3 - \boldsymbol{z}_2)\|_2^2\right| \\ &\leq_{(a)} \frac{1}{4}\max\begin{Bmatrix}\left|\left(1+\delta_{r,2M+s,\mu,l}\right)\|\boldsymbol{z}_3 + \boldsymbol{z}_2\|_2^2 - \left(1-\delta_{r,2M+s,\mu,l}\right)\|\boldsymbol{z}_3 - \boldsymbol{z}_2\|_2^2\right|, \\ \left|\left(1+\delta_{r,M+2s,\mu,l}\right)\|\boldsymbol{z}_3 - \boldsymbol{z}_2\|_2^2 - \left(1-\delta_{r,2M+s,\mu,l}\right)\|\boldsymbol{z}_3 + \boldsymbol{z}_2\|_2^2\right|\end{Bmatrix} \\ &= \left|\delta_{r,2M+s,\mu,l} + \langle \boldsymbol{z}_3, \boldsymbol{z}_2\rangle\right| \\ &\leq_{(b)} \delta_{r,2M+s,\mu,l},\end{aligned}$$

where inequality (a) follows the RIP property since $\boldsymbol{z}_3 + \boldsymbol{z}_2 \in MS_{r,2M+s,\mu,l}$ and $\boldsymbol{z}_3 - \boldsymbol{z}_2 \in MS_{r,2M+s,\mu,l}$; inequality (b) comes from that $\mathbf{T}_i^c \cap \mathbf{T}_j^c = \emptyset,\ \forall i \neq j$ and $\mathbf{T}_0 \cap \mathbf{T}_i^c = \emptyset, \forall i$.

Therefore,

$$\left|\langle \boldsymbol{A}\boldsymbol{R}^{\boldsymbol{m}}, \boldsymbol{A}\boldsymbol{R}^{\boldsymbol{a}}_{\mathbf{T}_j^c}\rangle\right| \leq \delta_{r,2M+s,\mu,l} \left\|\boldsymbol{R}^{\boldsymbol{a}}_{\mathbf{T}_0} + \boldsymbol{R}^{\boldsymbol{a}}_{\mathbf{T}_1^c}\right\|_2 \left\|\boldsymbol{R}^{\boldsymbol{a}}_{\mathbf{T}_j^c}\right\|_2. \tag{15}$$

Plugging Eq. (10), Eq. (14) and Eq. (15) into Eq. (13), we have

$$\begin{aligned}
&\left\|\boldsymbol{A}\left(\boldsymbol{R}^{\boldsymbol{a}}_{\mathbf{T}_0} + \boldsymbol{R}^{\boldsymbol{a}}_{\mathbf{T}_1^c}\right)\right\|_2^2 \\
&\leq \left\|\boldsymbol{R}^{\boldsymbol{a}}_{\mathbf{T}_0} + \boldsymbol{R}^{\boldsymbol{a}}_{\mathbf{T}_1^c}\right\|_2 \left(\sqrt{1+\delta_{r,M+s,\mu,l}}\epsilon_1 + \left(\delta_{r,M+s,\mu,l} + \tfrac{\eta_2}{1-\eta_2^2}\right)\|\boldsymbol{R}^{\boldsymbol{m}}\|_2 + \delta_{r,2M+s,\mu,l}\textstyle\sum_{j\geq 2}\left\|\boldsymbol{R}^{\boldsymbol{a}}_{\mathbf{T}_j^c}\right\|_2\right) \\
&\leq_{(a)} \left\|\boldsymbol{R}^{\boldsymbol{a}}_{\mathbf{T}_0} + \boldsymbol{R}^{\boldsymbol{a}}_{\mathbf{T}_1^c}\right\|_2 \left(\sqrt{1+\delta_{r,M+s,\mu,l}}\epsilon_1 + \left(\delta_{r,M+s,\mu,l} + \tfrac{\eta_2}{1-\eta_2^2}\right)\|\boldsymbol{R}^{\boldsymbol{m}}\|_2 + \delta_{r,2M+s,\mu,l}\sqrt{\tfrac{s}{M}}\gamma\left\|\boldsymbol{R}^{\boldsymbol{a}}_{\mathbf{T}_0}\right\|_2\right) \\
&\leq_{(b)} \left\|\boldsymbol{R}^{\boldsymbol{a}}_{\mathbf{T}_0} + \boldsymbol{R}^{\boldsymbol{a}}_{\mathbf{T}_1^c}\right\|_2 \left(\sqrt{1+\delta_{r,M+s,\mu,l}}\epsilon_1 + \left(\delta_{r,M+s,\mu,l} + \tfrac{\eta_2}{1-\eta_2^2}\right)\|\boldsymbol{R}^{\boldsymbol{m}}\|_2 + \delta_{r,2M+s,\mu,l}\sqrt{\tfrac{s}{M}}\gamma^2\left\|\boldsymbol{R}^{\boldsymbol{a}}_{\mathbf{T}_0} + \boldsymbol{R}^{\boldsymbol{a}}_{\mathbf{T}_1^c}\right\|\right),
\end{aligned}$$

where inequalities (a) and (b) follows the same argument as deriving Eq. (11).

According to the RIP property, we have

$$\begin{aligned}
&\left(1-\delta_{r,M+s,\mu,l}\right)\left\|\boldsymbol{R}^{\boldsymbol{a}}_{\mathbf{T}_0} + \boldsymbol{R}^{\boldsymbol{a}}_{\mathbf{T}_1^c}\right\|_2^2 \\
&\leq \left\|\boldsymbol{R}^{\boldsymbol{a}}_{\mathbf{T}_0} + \boldsymbol{R}^{\boldsymbol{a}}_{\mathbf{T}_1^c}\right\|_2 \left(\sqrt{1+\delta_{r,M+s,\mu,l}}\epsilon_1 + \left(\delta_{r,M+s,\mu,l} + \tfrac{\eta_2}{1-\eta_2^2}\right)\|\boldsymbol{R}^{\boldsymbol{m}}\|_2 + \delta_{r,2M+s,\mu,l}\sqrt{\tfrac{s}{M}}\gamma^2\left\|\boldsymbol{R}^{\boldsymbol{a}}_{\mathbf{T}_0}\right\|_2\right).
\end{aligned}$$

Consequently, we have

$$\begin{aligned}
&\left\|\boldsymbol{R}^{\boldsymbol{a}}_{\mathbf{T}_0} + \boldsymbol{R}^{\boldsymbol{a}}_{\mathbf{T}_1^c}\right\|_2 \\
&\leq \frac{\left(\sqrt{1+\delta_{r,M+s,\mu,l}}\epsilon_1 + \left(\delta_{r,M+s,\mu,l} + \dfrac{\eta_2}{1-\eta_2^2}\right)\|\boldsymbol{R}^{\boldsymbol{m}}\|_2 + \delta_{r,2M+s,\mu,l}\sqrt{\dfrac{s}{M}}\gamma^2\left\|\boldsymbol{R}^{\boldsymbol{a}}_{\mathbf{T}_0} + \boldsymbol{R}^{\boldsymbol{a}}_{\mathbf{T}_1^c}\right\|\right)}{\left(1-\delta_{r,M+s,\mu,l}\right)}
\end{aligned} \tag{16}$$

Notice that Eq. (11) and Eq. (16) still hold if we relax $\delta_{r,M+s,\mu,l}, \delta_{r,s,\mu,l}$ to $\delta_{r,2M+s,\mu,l}$. For simplicity, here we replace $\delta_{r,M+s,\mu,l}, \delta_{r,s,\mu,l}$ with $\delta_{r,2M+s,\mu,l}$ in the following derivation.

Plugging Eq. (11) into Eq. (16), and let $x \equiv \left\|\boldsymbol{R}^{\boldsymbol{a}}_{\mathbf{T}_0} + \boldsymbol{R}^{\boldsymbol{a}}_{\mathbf{T}_1^c}\right\|_2, y \equiv \|\boldsymbol{R}^{\boldsymbol{m}}\|_2$, we have

$$\left(D_1 - \frac{B_1 B_2}{D_2} - C_1\right)x \leq A_1\epsilon_1 + \frac{B_1}{D_2}A_2\epsilon_1 \tag{17}$$

where

$$A_1 = \sqrt{1+\delta_{r,2M+s,\mu,l}}, B_1 = \left(\delta_{r,2M+s,\mu,l} + \frac{\eta_2}{1-\eta_2^2}\right), C_1 = \delta_{r,2M+s,\mu,l}\sqrt{\frac{s}{M}}\gamma^2, D_1 = \left(1-\delta_{r,2M+s,\mu,l}\right),$$

and

$$\begin{aligned}
A_2 = \sqrt{1+\delta_{r,2M+s,\mu,l}}, B_2 = \left(\delta_{r,2M+s,\mu,l} + \frac{\eta_1}{1-\eta_1^2}\right)\sqrt{\frac{s}{M}}\gamma^2 + \left(\delta_{r,2M+s,\mu,l} + \frac{\eta_2}{1-\eta_2^2}\right), D_2 \\
= \left(1-\delta_{r,2M+s,\mu,l}\right).
\end{aligned}$$

Here we require $D_1 - \frac{B_1 B_2}{D_2} - C_1 > 0$ in Eq. (17) and let $M = s$, which is

$$1 - \gamma^2\alpha_1\alpha_2 - {\alpha_2}^2 - \left((1+\alpha_1+\alpha_2)\gamma^2 + 2\alpha_2 + 2\right)\delta_{r,3s,\mu,l} > 0. \tag{18}$$

Let $a = (1+\alpha_1+\alpha_2)\gamma^2 + 2\alpha_2 + 2$ and $c = 1-\gamma^2\alpha_1\alpha_2 - {\alpha_2}^2$, where $\alpha_1 = \frac{\eta}{1-\eta^2}$, $\alpha_2 = \frac{\sqrt{2}\eta}{1-2\eta^2}$, $\gamma = \sqrt{\frac{1+\delta_{B_a,2s}}{1-\delta_{B_a,2s}}}$, $\eta = \sqrt{\frac{\mu rsl}{n}}$. If $c > 0$, then, there exist a $\delta_{r,3s,\mu,l} > 0$ such that Eq. (18) is valid. Then the denominator $-a\delta_{r,3s,\mu,l} + c > 0\ \forall\ \delta_{r,3s,\mu,l} \in (0, c/a)$. Consequently,

$$\left\| \boldsymbol{R}^{\boldsymbol{a}}_{\mathbf{T}_0} + \boldsymbol{R}^{\boldsymbol{a}}_{\mathbf{T}_1^c} \right\|_2 \leq \frac{(1+\alpha_2)\sqrt{1+\delta_{r,3s,\mu,l}}}{c - a\delta_{r,3s,\mu,l}}\epsilon_1.$$

Notice that from Eq. (6) and Eq. (12), we have

$$\sum_{j\geq 2} \left\| \boldsymbol{R}^{\boldsymbol{a}}_{\mathbf{T}_j^c} \right\|_2 \leq \gamma \left\| \boldsymbol{R}^{\boldsymbol{a}}_{\mathbf{T}_0} \right\|_2 \leq \gamma^2 \left\| \boldsymbol{R}^{\boldsymbol{a}}_{\mathbf{T}_0} + \boldsymbol{R}^{\boldsymbol{a}}_{\mathbf{T}_1^c} \right\|_2$$

Therefore, $\|\boldsymbol{R}^{\boldsymbol{a}}\|_2 \leq \left\| \boldsymbol{R}^{\boldsymbol{a}}_{\mathbf{T}_0} + \boldsymbol{R}^{\boldsymbol{a}}_{\mathbf{T}_1^c} \right\|_2 + \sum_{j\geq 2} \left\| \boldsymbol{R}^{\boldsymbol{a}}_{\mathbf{T}_j^c} \right\|_2 \leq C_a\epsilon_1$, where

$$C_a = \frac{(1+\gamma^2)(1+\alpha_2)\sqrt{1+\delta_{r,3s,\mu,l}}}{c - a\delta_{r,3s,\mu,l}}.$$

Similarly, we can bound $\|\boldsymbol{R}^{\boldsymbol{m}}\|_2$ as $\|\boldsymbol{R}^{\boldsymbol{m}}\|_2 \leq C_m\epsilon_1$, where

$$C_m = \frac{\sqrt{1+\delta_{r,3s,\mu,l}} + \left(\delta_{r,3s,\mu,l} + \frac{\gamma^2}{1+\gamma^2}\alpha_1 + \frac{1}{1+\gamma^2}\alpha_2\right)C_a}{\left(1-\delta_{r,3s,\mu,l}\right)}.$$

Appendix E. Proof of Lemma 6.1.

In this section, we will provide the proof for Lemma 6.1. By linearity of the measurement matrix $\boldsymbol{A}$, without loss of generality, it is enough to prove this lemma when $\|\boldsymbol{y}\|_2 = 1$. The proof mainly has two steps. First, the bounds for $\boldsymbol{\theta}_a$ and $\boldsymbol{\theta}$ are derived and a finite set of points to approximate the set $MS_{r,\mathbf{T},\mu,l}$ to any accuracy in norm 2 sense can be found. Then, the concentration inequality can be applied through a union bound. This is a common approach in compressive sensing literature (Baraniuk et al., 2008; Tanner & Vary, 2020).

To derive the upper bounds for $\boldsymbol{\theta}_a$ and $\boldsymbol{\theta}$, we first derive the upper bounds for the signal $\boldsymbol{m} = \boldsymbol{B}\boldsymbol{\theta}$ and $\boldsymbol{a} = \boldsymbol{B}_a\boldsymbol{\theta}_a$, which are given in the following lemma.

Lemma 6.2. *The smooth signal component* $\boldsymbol{m}$ *and sparse signal component* $\boldsymbol{a}$ *of the signal* $\boldsymbol{y}$ *in* $MS_{r,\boldsymbol{T},\mu,l}$ *with* $\mu < \frac{n}{2rsl}$ *can be bounded as follows*

$$\|\boldsymbol{m}\|_2 = \|\boldsymbol{B}\boldsymbol{\theta}\|_2 \leq \frac{\|\boldsymbol{y}\|_2}{\sqrt{1-\eta^2}}, \tag{19}$$

$$\|\boldsymbol{a}\|_2 = \|\boldsymbol{B}_a\boldsymbol{\theta}_a\|_2 \leq \frac{\|\boldsymbol{y}\|_2}{\sqrt{1-\eta^2}}, \tag{20}$$

where $\eta = \sqrt{\frac{\mu rsl}{n}}$. ■

The proof is presented in Appendix F.

According to the RIP for $\boldsymbol{B}_a$, we have

$$\sqrt{(1-\delta_{\boldsymbol{B}_a,s})}\|\boldsymbol{\theta}_a\|_2 \le \|\boldsymbol{B}_a\boldsymbol{\theta}_a\|_2 \le \sqrt{(1+\delta_{\boldsymbol{B}_a,s})}\|\boldsymbol{\theta}_a\|_2. \tag{21}$$

Combine Eq. (20) and Eq. (21), we have

$$\|\boldsymbol{\theta}_a\|_2 \le \frac{1}{\sqrt{(1-\delta_{\boldsymbol{B}_a,s})}}\|\boldsymbol{B}_a\boldsymbol{\theta}_a\|_2 \le \frac{\|\boldsymbol{y}\|_2}{\sqrt{(1-\delta_{\boldsymbol{B}_a,s})(1-\eta^2)}} = \frac{1}{\sqrt{(1-\delta_{\boldsymbol{B}_a,s})(1-\eta^2)}}.$$

Denote $\tau_0 = \frac{1}{\sqrt{(1-\delta_{\boldsymbol{B}_a,s})(1-\eta^2)}}$, we have $\|\boldsymbol{\theta}_a\|_2 \le \tau_0$. Recall that $\boldsymbol{y} = \boldsymbol{B\theta} + \boldsymbol{B}_a\boldsymbol{\theta}_a$, we have $\boldsymbol{\theta} = \boldsymbol{B}^\dagger(\boldsymbol{y} - \boldsymbol{B}_a\boldsymbol{\theta}_a)$,where $\boldsymbol{B}^\dagger = \left(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}\right)^{-1}\boldsymbol{B}^{\mathrm{T}}$. Therefore, according to triangle inequality and Cauchy–Schwarz inequality, we have

$$\|\boldsymbol{\theta}\|_2 \le \left\|\boldsymbol{B}^\dagger\right\|_2(\|\boldsymbol{y}\|_2 + \|\boldsymbol{B}_a\boldsymbol{\theta}_a\|_2) \le \left\|\boldsymbol{B}^\dagger\right\|_2\left(1+\frac{1}{\sqrt{1-\eta^2}}\right).$$

Denote $\tau_1 = \left\|\boldsymbol{B}^\dagger\right\|_2\left(1+\frac{1}{\sqrt{1-\eta^2}}\right)$, we have $\|\boldsymbol{\theta}\|_2 \le \tau_1$.

Since we have derived the bounds for $\boldsymbol{\theta}_a$ and $\boldsymbol{\theta}$, the covering number of $MS_{r,\mathbf{T},\mu,l}$ is given by the following lemma whose proof is in Appendix G.

Lemma 6.3. *There exists a set $\boldsymbol{Q} \in MS_{r,\boldsymbol{T},\mu,l}$, such that for all $\boldsymbol{y} \in MS_{r,\boldsymbol{T},\mu,l}$, with $\|\boldsymbol{y}\|_2 = 1$ we have $\min_{\boldsymbol{q}\in\boldsymbol{Q}}\|\boldsymbol{q}-\boldsymbol{y}\|_2 \le \frac{\delta}{4}$, and $|\boldsymbol{Q}| \le \left(\frac{24}{\delta}\tau_1\right)^r\left(\frac{24}{\delta}\tau_0\right)^s$, where $|\boldsymbol{Q}|$ is its cardinality.* ■

Next, we will prove the main result by applying the concentration inequality (Definition 3.5. (ii)) with union bound. Let $\epsilon = \delta/2$,

$$\left(1-\frac{\delta}{2}\right)\|\boldsymbol{q}\|_2^2 \le \|\boldsymbol{Aq}\|_2^2 \le \left(1+\frac{\delta}{2}\right)\|\boldsymbol{q}\|_2^2 \quad \forall \boldsymbol{q} \in \mathbf{Q}, \tag{22}$$

with probability greater than $1 - 2|\mathbf{Q}|e^{-pc_0(\delta/2)}$.

Since $\delta \in (0,1)$, we have $1-\frac{\delta}{2} \le \sqrt{1-\frac{\delta}{2}}$ and $\sqrt{1+\frac{\delta}{2}} \le 1+\frac{\delta}{2}$. Then Eq. (22) can be written as

$$\left(1-\frac{\delta}{2}\right)\|\boldsymbol{q}\|_2 \le \|\boldsymbol{Aq}\|_2 \le \left(1+\frac{\delta}{2}\right)\|\boldsymbol{q}\|_2 \quad \forall \boldsymbol{q} \in \mathbf{Q}, \tag{23}$$

with probability greater than $1 - 2|\mathbf{Q}|e^{-pc_0(\delta/2)}$.

By the triangle inequality, we have

$$\|\boldsymbol{Ay}\|_2 \le \|\boldsymbol{A}(\boldsymbol{y}-\boldsymbol{q})\|_2 + \|\boldsymbol{Aq}\|_2. \tag{24}$$

Define

$$U = \max_{\boldsymbol{y}\in MS_{r,\mathbf{T},\mu,l},\|\boldsymbol{y}\|_2=1}\|\boldsymbol{Ay}\|_2, \tag{25}$$

which is attainable since $MS_{r,\mathbf{T},\mu,l}$ is closed.

Combine Eq. (23) and Eq. (24), we have $\forall \boldsymbol{y} \in MS_{r,\mathbf{T},\mu,l}$, with $\|\boldsymbol{y}\|_2 = 1$, there exist a $\boldsymbol{q} \in MS_{r,\mathbf{T},\mu,l}$, such that $\|\boldsymbol{Ay}\|_2 \le \|\boldsymbol{A}(\boldsymbol{q}-\boldsymbol{y})\|_2 + \left(1+\frac{\delta}{2}\right)\|\boldsymbol{q}\|_2$, with probability greater than $1 - 2|\mathbf{Q}|e^{-pc_0(\delta/2)}$.

Since $\boldsymbol{q}-\boldsymbol{y} \in MS_{r,\mathbf{T},\mu,l}$, if $\boldsymbol{q}-\boldsymbol{y} = \boldsymbol{0}$, we have $\|\boldsymbol{Ay}\|_2 \le \left(1+\frac{\delta}{2}\right)\|\boldsymbol{q}\|_2 = 1+\frac{\delta}{2}$.

If $\boldsymbol{q}-\boldsymbol{y}\neq\boldsymbol{0}$, we have $\|\boldsymbol{A}\boldsymbol{y}\|_2 \leq \left\|\boldsymbol{A}\frac{(\boldsymbol{q}-\boldsymbol{y})}{\|\boldsymbol{q}-\boldsymbol{y}\|_2}\right\|_2 \|\boldsymbol{q}-\boldsymbol{y}\|_2 + \left(1+\frac{\delta}{2}\right)\|\boldsymbol{q}\|_2 \leq \frac{\delta}{4}U + 1 + \frac{\delta}{2}$.

Notice that the second inequality comes from Eq.(25) combined with Q being a $\frac{\delta}{4}$ covering of $MS_{r,\mathbf{T},\mu,l}$. In summary, we have $\|\boldsymbol{A}\boldsymbol{y}\|_2 \leq \frac{\delta}{4}U + 1 + \frac{\delta}{2}$.

Since U is attainable, according to Eq. (25), we have $U \leq \frac{\delta}{4}U + 1 + \frac{\delta}{2}$. Consequently, we have $U \leq 1 + \frac{3}{4-\delta}\delta \leq 1+\delta$, since $\delta < 1$. Therefore, $\|\boldsymbol{A}\boldsymbol{y}\|_2 \leq 1+\delta$, with probability greater than $1 - 2|\mathbf{Q}|e^{-pc_0(\delta/2)}$.

Similarly, we can prove that $\|\boldsymbol{A}\boldsymbol{y}\|_2 \geq 1-\delta$with probability greater than $1-2|\mathbf{Q}|e^{-pc_0(\delta/2)}$.

Finally, according to Lemma 6.3, we have that

$$1-2|\mathbf{Q}|e^{-pc_0\left(\frac{\delta}{2}\right)} \geq 1-2\left(\frac{24}{\delta}\tau_1\right)^r\left(\frac{24}{\delta}\tau_0\right)^s e^{-pc_0\left(\frac{\delta}{2}\right)}.$$

This finishes the proof. ∎

Appendix F. Proof of Lemma 6.2.

To prove the result, we first derive a nontrivial upper bound for the inner produce between $\boldsymbol{m}$ and $\boldsymbol{a}$. Let $\boldsymbol{B}=\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$ be the reduced SVD of $\boldsymbol{B}$, then

$$\begin{aligned}
\left|\boldsymbol{m}^{\mathrm{T}}\boldsymbol{a}\right| &= \left|\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{a}\right| = \left|\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{U}^{T}\boldsymbol{a}\right| = \left|\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{U}^{T}\sum_{i}^{n} a_i\boldsymbol{e}_i\right| = \left|\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{\Sigma}\sum_{i}^{n} a_i\boldsymbol{U}^{T}\boldsymbol{e}_i\right| \\
&\leq \left\|\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{\Sigma}\right\|_2 \left\|\sum_{i}^{n} a_i\boldsymbol{U}^{T}\boldsymbol{e}_i\right\|_2 \\
&\leq \left\|\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{\Sigma}\right\|_2 \sum_{i}^{n}|a_i|\left\|\boldsymbol{U}^{T}\boldsymbol{e}_i\right\|_2 \\
&\leq \left\|\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{\Sigma}\right\|_2 \sum_{i}^{n}|a_i| \max_{j\in\{1,\ldots r\}}\left\|\boldsymbol{U}^{T}\boldsymbol{e}_j\right\|_2 \\
&\leq_{(a)} \left\|\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{U}^{\mathrm{T}}\right\|_2 \|\boldsymbol{a}\|_1 \sqrt{\frac{\mu r}{n}} \\
&\leq_{(b)} \sqrt{\frac{\mu r s l}{n}}\|\boldsymbol{m}\|_2\|\boldsymbol{a}\|_2.
\end{aligned} \tag{26}$$

where inequality (a) follows $\left\|\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{\Sigma}\right\|_2 = \left\|\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{U}^{\mathrm{T}}\right\|_2$ since $\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}=\boldsymbol{I}$ and Definition 2.2. Inequality (b) follows from $\|\boldsymbol{a}\|_1 \leq \sqrt{ls}\|\boldsymbol{a}\|_2$.

Let $\eta = \sqrt{\frac{\mu r s l}{n}}$, since $\mu < \frac{n}{rsl}$, we have $\eta < 1$ and

$$\left|\boldsymbol{m}^{\mathrm{T}}\boldsymbol{a}\right| = \frac{\left|\|\boldsymbol{y}\|_2^2 - \|\boldsymbol{m}\|_2^2 - \|\boldsymbol{a}\|_2^2\right|}{2} \leq \eta\|\boldsymbol{m}\|_2\|\boldsymbol{a}\|_2.$$

Therefore, we have

$$\|\boldsymbol{m}\|_2^2 + \|\boldsymbol{a}\|_2^2 - \|\boldsymbol{y}\|_2^2 \le 2\eta\|\boldsymbol{m}\|_2\|\boldsymbol{a}\|_2.$$

By completing the square, we have

$$(\|\boldsymbol{m}\|_2 + \eta\|\boldsymbol{a}\|_2)^2 + (1-\eta^2)\|\boldsymbol{a}\|_2^2 - \|\boldsymbol{y}\|_2^2 \le 0.$$

Since $(\|\boldsymbol{m}\|_2 + \eta\|\boldsymbol{a}\|_2)^2 \ge 0$, we have

$$\|\boldsymbol{a}\|_2 \le \frac{1}{\sqrt{(1-\eta^2)}}\|\boldsymbol{y}\|_2.$$

Similarly, we can derive that

$$\|\boldsymbol{m}\|_2 \le \frac{1}{\sqrt{(1-\eta^2)}}\|\boldsymbol{y}\|_2.$$

This finishes the proof. ■

Appendix G. Proof of Lemma 6.3.

We first state results for the covering number of a set (Vershynin, 2018): The covering number of a smallest ϵ-net for a unit l_2 norm ball in d dimensional space is $(3/\epsilon)^d$.

Let $\mathbf{M} = \{\boldsymbol{m} \in \mathbb{R}^n | \boldsymbol{m} = \boldsymbol{B\theta}, \boldsymbol{\theta} \in \mathbb{R}^r, \|\boldsymbol{\theta}\|_2 \le \tau_1, \mu(\boldsymbol{B}) = \mu\}$ and $\mathbf{S} = \{\boldsymbol{a} \in \mathbb{R}^n | \boldsymbol{a} = \boldsymbol{B}_a\boldsymbol{\theta}_a, \boldsymbol{\theta}_a \in \mathbf{T}, \|\boldsymbol{\theta}_a\|_2 \le \tau_0, l(\boldsymbol{B}_a) = l\}$. There exists two finite $\frac{\delta}{8}$ covering sets of $\mathbf{M}$ and $\mathbf{S}$, which are $\mathbf{Q_M} \subseteq \mathbf{M}$ and $\mathbf{Q_S} \subseteq \mathbf{S}$.

For all $\boldsymbol{q}_{\mathbf{M}} \in \mathbf{Q_M}$, and for all $\boldsymbol{m} \in \mathbf{M}$, we have,

$$\min_{\boldsymbol{q}_{\mathbf{M}} \in \mathbf{Q_M}} \|\boldsymbol{m} - \boldsymbol{q}_{\mathbf{M}}\|_2 \le \frac{\delta}{8};$$

For all $\boldsymbol{q}_{\mathbf{S}} \in \mathbf{Q_s}$, and for all $\boldsymbol{a} \in \mathbf{S}$, we have,

$$\min_{\boldsymbol{q}_{\mathbf{S}} \in \mathbf{Q_s}} \|\boldsymbol{a} - \boldsymbol{q}_{\mathbf{S}}\|_2 \le \frac{\delta}{8}.$$

According to (i) and (ii), we have $|\mathbf{Q_M}| \le \left(\frac{24}{\delta}\tau_1\right)^r$ and $|\mathbf{Q_s}| \le \left(\frac{24}{\delta}\tau_0\right)^s$.

Define $\mathbf{Q_{MS}} = \{\boldsymbol{q}_{\mathbf{M}} + \boldsymbol{q}_{\mathbf{S}} | \boldsymbol{q}_{\mathbf{M}} \in \mathbf{Q_M}, \boldsymbol{q}_{\mathbf{S}} \in \mathbf{Q_S}\} \subseteq MS_{r,\mathbf{T},\mu,l}$. Then $\forall \boldsymbol{y} \in MS_{r,\mathbf{T},\mu,l}$, there exists a pair $\boldsymbol{q}_{\mathbf{MS}} = \boldsymbol{q}_{\mathbf{M}} + \boldsymbol{q}_{\mathbf{S}} \in MS_{r,\mathbf{T},\mu,l}$, such that

$$\|\boldsymbol{q}_{\mathbf{MS}} - \boldsymbol{y}\|_2 = \|\boldsymbol{q}_{\mathbf{M}} - \boldsymbol{m} + \boldsymbol{q}_{\mathbf{S}} - \boldsymbol{a}\|_2 \le \|\boldsymbol{q}_{\mathbf{M}} - \boldsymbol{m}\|_2 + \|\boldsymbol{q}_{\mathbf{S}} - \boldsymbol{a}\|_2 \le \frac{\delta}{4}.$$

Therefore, $\mathbf{Q_{MS}}$ is a $\delta/4$ covering of $MS_{r,\mathbf{T},\mu,l}$ and $|\mathbf{Q_{MS}}| \le \left(\frac{24}{\delta}\tau_1\right)^r \left(\frac{24}{\delta}\tau_0\right)^s$. This finishes the proof. ■

References

Baraniuk, R., Davenport, M., DeVore, R., & Wakin, M. (2008). A simple proof of the restricted isometry property for random matrices. *Constructive Approximation*, *28*(3), 253-263.

Candes, E. J., Romberg, J. K., & Tao, T. (2006). Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, *59*(8), 1207-1223.

Candes, E. J., & Tao, T. (2005). Decoding by linear programming. *IEEE Transactions on Information Theory*, *51*(12), 4203-4215.

Tanner, J., & Vary, S. (2020). Compressed sensing of low-rank plus sparse matrices. *arXiv preprint arXiv:2007.09457*.

Vershynin, R. (2018). High-dimensional probability: An introduction with applications in data science (Vol. 47). Cambridge University Press, Cambridge.