Some details of high-order spectral method

Chanwoo Lee, July 19, 2021

Let us consider $\mathcal{Y} = \Theta + \mathcal{E} \in \mathbb{R}^{d_1 \times \cdots \times d_m}$ where $\mathcal{E}$ follows i.i.d. sub-Gaussian noise with $\sigma^2 = 1$ without loss of generality and $\Theta = \mathcal{X} + \mathcal{X}_\perp$ with $\text{rank}(\mathcal{X}) = (\sqrt{d_1}, \ldots, \sqrt{d_m})$. For each $k = 1, ..., n$, denote

$$X_k = \mathcal{M}_k(\mathcal{X}), \quad X_{k,\perp} = \mathcal{M}_k(\mathcal{X}_\perp), \quad E_k = \mathcal{M}_k(\mathcal{E}), \quad Y_k = \mathcal{M}_k(\mathcal{Y}),$$

and define $Z_k = X_{k,\perp} + E_k$. We consider the high-order spectral method, where we estimate the signal tensor Θ from

$$\begin{aligned}
\tilde{U}_k &= \text{SVD}_{r_k}(Y_k) \\
\hat{U}_k &= \text{SVD}_{r_k}\left(\mathcal{M}_k\left(\mathcal{Y} \times_1 \tilde{U}_1^T \times \cdots \times_{k-1} \tilde{U}_{k-1}^T \times_k \tilde{U}_{k+1}^T \times \cdots \times_m \tilde{U}_m^T\right)\right) \\
\hat{\Theta} &= \mathcal{Y} \times_1 (\hat{U}_1\hat{U}_1^T) \times \cdots \times_m (\hat{U}_m\hat{U}_m^T),
\end{aligned} \tag{1}$$

What's the implication for smooth tensor?
Consider order-3 tensor with equal dimension.
By approximation theorem, smooth tensor has the following properties:
spectral norm of (X_perp) <= F-norm of (X_prep) < = sqrt(d^3/d) = d

where $r_k = \sqrt{d_k}$ for all $k \in [m]$.

Theorem 0.1 (Estimation of high-order spectral algorithm)**.** Suppose that $\|X_\perp\|_{\text{sp}} \leq \sqrt{\bar{d}}$. Then, with probability at least $1 - C\exp(-c\underline{d})$, $\hat{\Theta}$ defined according to (1) satisfies,

$$\|\hat{\Theta} - \Theta\|_F^2 \lesssim \bar{d}^2 + (\bar{d}d_*)^{1/2}.$$

Proof. We start by introducing several notations and assumptions. Denote $U_k = \text{SVD}_{r_k}(X_k)$ and $\tilde{U}_k = \text{SVD}_{r_k}(Y_k)$. For some constant $C_0 > 0$ which will be specified later, define

$$r_k' = \max\{r' \in \{0, \ldots, d_k\} \colon \sigma_{r'}(X_k) \geq C_0\cancel{(d_*^{1/4} \vee \bar{d}^{1/2})}\}.$$

use threshold d for smooth tensor
(see page 3, green comments for reason)

We set $r_k' = 0$ if $\sigma_1(X_k) < C_0(d_*^{1/4} \vee \bar{d}^{1/2})$ We use U_k to denote the leading r_k' singular vectors of U_k and use V_k to denote the rest $r_k - r_k'$ singular vectors and thus U_k can be written as $[U_k, V_k]$. We next define

$$X_k' = X_k\left(\mathbb{P}_{U_{k+1}'} \otimes \cdots \otimes \mathbb{P}_{U_m'} \otimes \cdots \otimes \mathbb{P}_{U_{k-1}'}\right),$$

where $\mathbb{P}_U = UU^T$ for any orthonormal matrix $U \in \mathbb{R}^{d \times r}$. We also denote

$$\begin{aligned}
\bar{X}_k &= X_k\left(\tilde{U}_{k+1} \otimes \cdots \otimes \tilde{U}_m \otimes \cdots \otimes \tilde{U}_{k-1}\right), \\
\bar{X}_{k,\perp} &= X_{k,\perp}\left(\tilde{U}_{k+1} \otimes \cdots \otimes \tilde{U}_m \otimes \cdots \otimes \tilde{U}_{k-1}\right), \\
\bar{Y}_k &= Y_k\left(\tilde{U}_{k+1} \otimes \cdots \otimes \tilde{U}_m \otimes \cdots \otimes \tilde{U}_{k-1}\right), \\
\bar{E}_k &= E_k\left(\tilde{U}_{k+1} \otimes \cdots \otimes \tilde{U}_m \otimes \cdots \otimes \tilde{U}_{k-1}\right),
\end{aligned}$$

and define $\bar{Z}_k = \bar{X}_{k,\perp} + \bar{E}_k$.

Now we bound

$$\|\hat{\Theta} - \Theta\|_F \leq \underbrace{\|\Theta \times_1 (\hat{U}_1\hat{U}_1^T) \times \cdots \times_m (\hat{U}_m\hat{U}_m^T) - \mathcal{X}\|_F}_{(*)} + \underbrace{\|\mathcal{E} \times_1 \hat{U}_1^T \times \cdots \times_m \hat{U}_m^T\|_F}_{(**)}.$$

To bound $(*)$, we have

$$
\begin{aligned}
(*) &\le \sum_{k\in[m]} \|(I-\hat U_k \hat U_k^T)\Theta_k\|_F \\
&\le \sum_{k\in[m]} \left(\|(I-\hat U_k \hat U_k^T) X_k\|_F + \|(I-\hat U_k \hat U_k^T) X_{k,\perp}\|_F \right) \\
&\le \sum_{k\in[m]} \left(\|\hat U_{k,\perp}^T X_k\|_F + \|X_{k,\perp}\|_F \right) \\
&\le \sum_{k\in[m]} \left(\|\hat U_{k,\perp}^T X_k'\|_F + \|X_k - X_k'\|_F + \|X_{k,\perp}\|_F \right). \qquad (2)
\end{aligned}
$$

Therefore, it suffices to bound $\|X_k - X_k'\|_F$ and $\|\hat U_{k,\perp}^T X_k'\|_F$.

1. Bound of $\|X_k - X_k'\|_F$: For notation simplicity, we focus on $k=1$, while the analysis for other modes can be similarly carried on.

$$
\begin{aligned}
\|X_1 - X_1'\|_F &\le \|X_1 \left((\mathbb{P}_{U_2'} + \mathbb{P}_{V_2'}) \otimes \cdots \otimes (\mathbb{P}_{U_m'} + \mathbb{P}_{U_m'}) - \mathbb{P}_{U_2'} \otimes \cdots \otimes \mathbb{P}_{U_m'} \right)\|_F \\
&\le \sum_{k=2}^m \|V_k'^T X_k\|_F \\
&\le \sum_{k=2}^m \sqrt{r_k - r_k'}\,\sigma_{r_k'+1}(X_k) \\
&\le \sum_{k=2}^m c_0 \cancel{\sqrt{r_k}\, d_*^{1/4}}. \qquad (3)
\end{aligned}
$$

sqrt(r)*d.
d is due to the smooth tensor; see green comments at page 3

2. Bound of $\|\hat U_{k,\perp}^T X_k'\|_F$: Notice the following two inequalitoes,

$$
\|\hat U_{k,\perp}^T X_k' (\tilde U_{k+1} \otimes \cdots \otimes \tilde U_m \otimes \tilde U_1 \otimes \cdots \otimes \tilde U_{k-1}\|_F \le \|\hat U_{k,\perp}^T \bar X_k\|_F + \|X_k - X_k'\|_F \qquad (4)
$$

$$
\begin{aligned}
&\|\hat U_{k,\perp}^T X_k' (\tilde U_{k+1} \otimes \cdots \otimes \tilde U_m \otimes \tilde U_1 \otimes \cdots \otimes \tilde U_{k-1}\|_F \\
&\quad = \|\hat U_{k,\perp}^T X_k' (\mathbb{P}_{U_{k+1}'} \tilde U_{k+1} \otimes \cdots \otimes \mathbb{P}_{U_m'} \tilde U_m \otimes \mathbb{P}_{U_1'} \tilde U_1 \otimes \cdots \otimes \mathbb{P}_{U_{k-1}'} \tilde U_{k-1}\|_F \\
&\quad = \|\hat U_{k,\perp}^T X_k'\|_F \prod_{\ell\neq k} \sigma_{r_\ell'}(U_\ell^T \tilde U_\ell) \\
&\quad \ge \|\hat U_{k,\perp}^T X_k'\|_F \prod_{\ell\neq k} \sqrt{1 - \|\tilde U_{\ell,\perp} U_\ell'\|_{\mathrm{sp}}^2} \qquad (5)
\end{aligned}
$$

Combining (4) and (5) yields,

$$
\|\hat U_{k,\perp}^T X_k'\|_F \prod_{\ell\neq k} \sqrt{1 - \|\tilde U_{\ell,\perp} U_\ell'\|_{\mathrm{sp}}^2} \le \|\hat U_{k,\perp}^T \bar X_k\|_F + \|X_k - X_k'\|_F. \qquad (6)
$$

Now, we bound $\|\hat U_{k,\perp}^T \bar X_k\|_F$ and $\|\tilde U_{\ell,\perp} U_\ell'\|_{\mathrm{sp}}$ to obtain upper bound of $\|\hat U_{k,\perp}^T X_k'\|_F$.

We have an upper bound for $\|\hat{U}_{k,\perp}^T \bar{X}_k\|_F$, combining of Lemma 0.2 and the fact that $\bar{Y}_k = \bar{X}_k + \bar{Z}_k$,

$$\begin{aligned}\|\hat{U}_{k,\perp}^T \bar{X}_k\|_F &\leq 2\sqrt{r_k}\|\bar{Z}\|_{\mathrm{sp}} \\ &\leq 2\sqrt{r_k}\left(\|\bar{X}_{k,\perp}\|_{\mathrm{sp}} + \|\bar{E}_k\|_{\mathrm{sp}}\right) \\ &\leq 2\sqrt{r_k}\left(\|X_{k,\perp}\|_{\mathrm{sp}} + \|\bar{E}_k\|_{\mathrm{sp}}\right) \\ &\lesssim \sqrt{r_k \bar{d}} + \sqrt{r_*} + \sum_{\ell\in[m]} \sqrt{r_\ell \bar{r} d_\ell}. \end{aligned} \tag{7}$$

where the last line uses the condition $\|X_\perp\|_{\mathrm{sp}} \leq \sqrt{\bar{d}}$, the definition of $\bar{E}_k$ and Lemma 0.1.

By Lemma 0.3, we bound $\|\tilde{U}_{k,\perp} U_k'\|_{\mathrm{sp}}$ with probability at least $1 - C\exp(-c\underline{d})$, for each $k \in [m]$,

$$\begin{aligned}\|\tilde{U}_{k,\perp} U_k'\|_{\mathrm{sp}} &\leq C\left(\frac{\sqrt{d_k} + \|X_{k,\perp}\|_{\mathrm{sp}}}{\sigma_{r'}(X)} + \frac{\sqrt{d_*} + \|X_{k,\perp}\|_{\mathrm{sp}}^2}{\sigma_{r'}^2(X)}\right) \\ &\leq \frac{C}{C_0}\left(\frac{\sqrt{d_k} + \sqrt{\bar{d}}}{\sqrt{d_*}} + \frac{\sqrt{d_*} + \bar{d}}{\sqrt{d_*} \vee \bar{d}}\right) \\ &\leq \frac{1}{\sqrt{2}} \end{aligned} \tag{8}$$

numerator is d^2, because of smooth tensor.

In order for the ratio to be a constant, the denominator has to be ~ d^2.

Therefore, we set the threshold for the denominator: sigma_r'(X)~d in page 1.

for sufficiently large $C_0 \geq 15$ where 15 is set to satisfy the condition of Lemma 0.3.

Finally, plugging (3),(7), and (8) into (6) yields,

$$\begin{aligned}\|\hat{U}_{k,\perp}^T X_k'\|_F &\leq 2^{\frac{m-1}{2}}\left(\|\hat{U}_{k,\perp}^T \bar{X}_k\|_F + \|X_k - X_k'\|_F\right) \\ &\lesssim r_*^{1/2} + \bar{r}\bar{d}^{1/2} + \bar{r}^{1/2} d_*^{1/4}. \end{aligned} \tag{9}$$

Applying (3) and (9) to (2) proves

$$(*) \lesssim r_*^{1/2} + \bar{r}\bar{d}^{1/2} + \bar{r}^{1/2} d_*^{1/4}.$$

Notice that $(**)$ term is bounded by $C(\sqrt{r_*} + \sum_{\ell\in[m]} \sqrt{d_\ell r_\ell})$ by Lemma 0.1 with probability at least $1 - \exp(-c\underline{d})$. Combining upper bound of $(*)$ and $(**)$, we finally obtain

for smooth tensor, F-norm error becomes d^{5/2}

$$\|\hat{\Theta} - \Theta\|_F \lesssim r_*^{1/2} + \bar{r}\bar{d}^{1/2} + \bar{r}^{1/2} d_*^{1/4}. \tag{10}$$

—> || ||_F^2/d^3 ~ d^{-1/2}

Even worse than square spectral method.

Plugging $r_k = \sqrt{d_k}$ for all $k \in [m]$ into (10) completes the proof. □

Lemma 0.1 (Lemma 8 in [2]). Let $E \in \mathbb{R}^{d_1\times\cdots\times d_m}$ be a noise tensor whose each entry has independent mean-zero sub-Gaussian distribution with $\sigma = 1$ without loss of generality. Fix $U_k^* \in \mathbb{O}_{d_k, r_k}$. Then with probability at least $1 - \exp(-c\underline{d})$, the following holds.

$$\begin{aligned}\|E_k\left(U_{k+1}^* \otimes \cdots \otimes U_m^* \otimes U_1^* \otimes \cdots \otimes U_{k-1}^*\right)\|_{\mathrm{sp}} &\leq C(\sqrt{d_k} + \sqrt{r_{-k}}), \\ \|E_k\left(U_{k+1}^* \otimes \cdots \otimes U_m^* \otimes U_1^* \otimes \cdots \otimes U_{k-1}^*\right)\|_F &\leq C\sqrt{d_k r_{-k}}, \end{aligned}$$

$$\sup_{\substack{U_\ell \in \mathbb{O}_{d_\ell, r_\ell} \\ \ell \neq [m]}} \|E_k \left(U_{k+1} \otimes \cdots \otimes U_m \otimes U_1 \otimes \cdots \otimes U_{k-1}\right)\|_{\text{sp}} \leq C(\sqrt{d_k} + \sqrt{r_{-k}} + \sum_{\ell \neq k} \sqrt{d_\ell r_\ell}),$$

$$\sup_{\substack{U_\ell \in \mathbb{O}_{d_\ell, r_\ell} \\ \ell \neq [m]}} \|E_k \left(U_{k+1} \otimes \cdots \otimes U_m \otimes U_1 \otimes \cdots \otimes U_{k-1}\right)\|_F \leq C(\sqrt{d_k r_{-k}} + \sum_{\ell \neq k} \sqrt{d_\ell r_\ell}),$$

$$\sup_{\substack{U_\ell \in \mathbb{O}_{d_\ell, r_\ell} \\ \ell \neq [m]}} \|\mathcal{E} \times_1 U_1^T \times \cdots \times_m U_m^T\|_F \leq C(\sqrt{r_*} + \sum_{\ell \in [m]} \sqrt{d_\ell r_\ell})$$

Lemma 0.2 (Projection bound of perturbation)**.** Suppose $X, E \in \mathbb{R}^{m \times n}$ and $\text{rank}(X) = r$. Let $U \in \mathbb{O}_{m,r}$ be the leading r singular vectors of $Y = X + E$. Then,

$$\|(I - UU^T)X\|_{\text{sp}} \leq 2\|E\|_{\text{sp}}$$
$$\|(I - UU^T)X\|_F \leq \min\left(2\sqrt{r}\|E\|_{\text{sp}}, 2\|E\|_F\right).$$

Proof. For matrix norm bound we have,

$$\begin{aligned}
\|(I - UU^T)X\|_{\text{sp}} &\leq \|(I - UU^T)Y\|_{\text{sp}} + \|E\|_{\text{sp}} \\
&\leq \sigma_{r+1}(Y) + \|E\|_{\text{sp}} \\
&\leq \min_{Z \in \mathbb{R}^{m \times n}:\, \text{rank}(Z) \leq r} \|Y - Z\|_{\text{sp}} + \|E\|_{\text{sp}} \\
&\leq \|Y - X\|_{\text{sp}} + \|E\|_{\text{sp}} \\
&\leq 2\|E\|_{\text{sp}}. \qquad (11)
\end{aligned}$$

Similarly we bound Frobenius norm,

$$\begin{aligned}
\|(I - UU^T)X\|_F &\leq \|(I - UU^T)Y\|_F + \|E\|_F \\
&\leq \sqrt{\sum_{i=r+1}^{m \wedge n} \sigma_i^2(Y)} + \|E\|_F \\
&\leq \min_{Z \in \mathbb{R}^{m \times n}:\, \text{rank}(Z) \leq r} \|Y - Z\|_F + \|E\|_F \\
&\leq \|Y - X\|_F + \|E\|_F \\
&\leq 2\|E\|_F.
\end{aligned}$$

In addition, direct application of (11) yields,

$$\|(I - UU^T)X\|_F \leq 2\sqrt{r}\|E\|_{\text{sp}}.$$

□

Lemma 0.3 (Perturbation Bound on Subspaces of Different Dimensions)**.** Consider the signal plus noise model,

$$Y = X + X_\perp + E \in \mathbb{R}^{d_1 \times d_2},$$

where X is a signal matrix such that $\text{rank}(X) = r$, $X_\perp$ is a perturbation, and E is a noise matrix with i.i.d. standard sub-Gaussian entries. Define

$$r' := \max\{r' \in \{0,1,\dots,r\} \colon \sigma_{r'}(X) \geq \max(\sqrt{3d_1}, (8+6\sqrt{2})\|X_\perp\|_{\mathrm{sp}})\}. \tag{12}$$

We denote

$$\hat{U}_r = \mathrm{SVD}_r(Y), \quad U_{r'} = \mathrm{SVD}_{r'}(X).$$

Then with probability at least $1-\exp(-cd_1 \wedge d_2)$,

$$\|\hat{U}_{r\perp} U_{r'}\|_{\mathrm{sp}} \leq C\left(\frac{\sqrt{d_1}+\|X_\perp\|_{\mathrm{sp}}}{\sigma_{r'}(X)} + \frac{\sqrt{d_1 d_2}+\|X_\perp\|_{\mathrm{sp}}^2}{\sigma_{r'}^2(X)}\right).$$

Proof. Since each entry of E comes from i.i.d. sub-Gaussian distribution, the model does no change by left multiplying some orthogonal matrix in $\mathbb{O}_{d_1,d_1}$ on $X, X_\perp$ and Z simultaneously. Thus we can assume that $\mathrm{SVD}_r(X) = [I_r O]^T$ without loss of generality. We write,

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}, \quad X_\perp = \begin{pmatrix} X_{1,\perp} \\ X_{2,\perp} \end{pmatrix}, \quad X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \quad E = \begin{pmatrix} E_1 \\ E_2 \end{pmatrix},$$

where $Y_1, X_{1,\perp}, X_1, E_1 \in \mathbb{R}^{r' \times p_2}$ and $Y_2, X_{2,\perp}, X_2, E_2 \in \mathbb{R}^{(p_1-r')\times p_2}$. We calculate the SVD of $Y_1 = \bar{U}\bar{\Sigma}\bar{V}^T$ for $\bar{U} \in \mathcal{O}_{r',r'}, \bar{V} \in \mathcal{O}_{p_2,r'}$. Then, by the proof of Lemma 2 in [2], we have

$$\|\hat{U}_{r\perp} U_{r'}\|_{\mathrm{sp}} \leq \frac{\sigma_{r'}(Y_1)\|Y_2\bar{V}\|_{\mathrm{sp}}}{\sigma_{r'}^2(Y_1) - \sigma_{r+1}^2(Y)}. \tag{13}$$

Therefore, it suffices to provide the probabilistic bounds of $\sigma_{r'}^2(Y_1) - \sigma_{r+1}^2(Y)$, $\sigma_{r'}(Y_1)$, and $\|Y_2\bar{V}\|_{\mathrm{sp}}$.

1. Bound of $\sigma_{r'}^2(Y_1) - \sigma_{r+1}^2(Y)$: Define $Y' = X + E$, where the perturbation matrix $X_\perp$ is removed. By [1, Lemma 9], for all $x > 0$, we have

$$\mathbb{P}\left(\sigma_{r'}^2(Y_1') \leq (\sigma_{r'}^2(X_1)+d_2)(1-x)\right) \leq C\exp(Cr - c(\sigma_{r'}(X_1)+d_2)x^2 \wedge x),$$
$$\mathbb{P}\left(\sigma_{r+1}^2(Y') \leq d_2(1+x)\right) \leq C\exp(Cd_1 - cd_2x^2 \wedge x).$$

By setting x as $\frac{\sigma_{r'}^2(X_1)}{3(\sigma_{r'}^2(X_1)+d_2)}$ and $\frac{\sigma_{r'}^2(X_1)}{3d_2}$ respectively, we obtain

$$\sigma_{r'}(Y_1') \geq \sqrt{\frac{2\sigma_{r'}^2(X_1)}{3}+d_2}, \quad \sigma_{r+1}(Y') \geq \sqrt{\frac{\sigma_{r'}^2(X_1)}{3}+d_2},$$

with probability at least 1-$C\exp(-cd_1 \wedge d_2)$. Since $Y = Y' + X_\perp$, applying Weyl's inequality yields

$$\sigma_{r'}(Y_1) \geq \sqrt{\frac{2\sigma_{r'}^2(X_1)}{3}+d_2} - \|X_{1,\perp}\|_{\mathrm{sp}}, \quad \sigma_{r+1}(Y) \geq \sqrt{\frac{\sigma_{r'}^2(X_1)}{3}+d_2} + \|X_{1,\perp}\|_{\mathrm{sp}}. \tag{14}$$

Therefore, we obtain the following inequality from (14),

$$\begin{aligned}\sigma_{r'}^2(Y_1) - \sigma_{r+1}^2(Y) &\geq \frac{\sigma_{r'}^2(X_1)}{3} - 4\|X_{1,\perp}\|_{\mathrm{sp}}\sqrt{\frac{2\sigma_{r'}^2(X_1)}{3} + d_2} - 2\|X_{1,\perp}\|_{\mathrm{sp}} \\ &\geq \frac{\sigma_{r'}^2(X_1)}{3} - 4\sigma_{r'}(X_1)\|X_\perp\|_{\mathrm{sp}} - 2\|X_\perp\|_{\mathrm{sp}} \\ &\geq \frac{\sigma_{r'}^2(X_1)}{12}, \end{aligned} \tag{15}$$

where the second inequality uses the fact $\|X_{1,\perp}\|_{\mathrm{sp}} \vee \|X_{2,\perp}\|_{\mathrm{sp}} \leq \|X_\perp\|_{\mathrm{sp}}$ with the condition $\sigma_{r'}(X) \geq \sqrt{3d_2}$ while the last inequality holds with the condition $\sigma_{r'}(X) \geq (8+6\sqrt{2})\|X_\perp\|_{\mathrm{sp}}$. Notice that the definition of r' in (12) makes (15) true.

2. Bound of $\sigma_{r'}(Y_1)$: By Weyl's inequality, we easily obtain the upper bound of $\sigma_{r'}(Y_1)$,

$$\sigma_{r'}(Y_1) \leq \sigma_{r'}(X_1) + \|X_{1,\perp}\|_{\mathrm{sp}} + \|E_1\|_{\mathrm{sp}} \leq \sigma_{r'}(X_1) + \|X_\perp\|_{\mathrm{sp}} + \sqrt{d_2}, \tag{16}$$

where the last inequality holds with probability at least $1 - C\exp(-cd_2)$ by the concentration inequality of $\|E_1\|_{\mathrm{sp}}$.

3. Bound of $\|Y_2\bar{V}\|_{\mathrm{sp}}$: Notice that

$$\|Y_2\bar{V}\|_{\mathrm{sp}} \leq \|X_2\bar{V}\|_{\mathrm{sp}} + \|E_2\bar{V}\|_{\mathrm{sp}} + \|X_{2,\perp}\|_{\mathrm{sp}} \tag{17}$$

Note that $Y_1 = \bar{U}\bar{\Sigma}\bar{V}$, we have,

$$\|\bar{V}^T X_2^T\|_{\mathrm{sp}} \leq \frac{\|Y_1 X_2^T\|_{\mathrm{sp}}}{\sigma_{r'}(Y_1)} \leq \frac{\|(X_1 + X_{1,\perp} + E_1)X_2^T\|_{\mathrm{sp}}}{\sigma_{r'}(Y_1)} \tag{18}$$

Recall that we assume that $\mathrm{SVD}_r(X) = [I_r, O]^T$. Therefore, X can be written as,

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} I_{r'} & \\ & I_{r-r'} \end{pmatrix} \begin{pmatrix} \Sigma_1 & \\ & \Sigma_2 \end{pmatrix} \begin{pmatrix} V_1^T \\ V_2^T \end{pmatrix} = \begin{pmatrix} \Sigma_1 V_1^T \\ \Sigma_2 V_2^T \\ O_{p_1 - r} \end{pmatrix}.$$

From this expression, we have $X_1 = \Sigma_1 V_1^T$ and $X_1 X_2^T = 0$. Therefore, (18) becomes

$$\|\bar{V}^T X_2^T\|_{\mathrm{sp}} \leq \frac{\|(X_{1,\perp} + E_1)V_2\Sigma_2\|_{\mathrm{sp}}}{\sigma_{r'}(Y_1)} \leq \frac{\sigma_{r'+1}(X)}{\sigma_{r'}(Y_1)} \left(\|X_{1,\perp}\|_{\mathrm{sp}} + \|E_1 V_2\|_{\mathrm{sp}}\right).$$

Since $E_1 V_2$ is a $r' \times (r - r')$ standard Gaussian random matrix, we have $\|E_1 V_2\|_{\mathrm{sp}} \leq \sqrt{d_1}$ with probability at least $1 - C\exp(-cd_1)$. On the other hand, since $\bar{V}$ is the leading r' right singular vectors of $Y_1 = X_1 + X_{1,\perp} + E_1$, we know that $\bar{V}$ is independent of Y_2 and $E_2 V$ is then a $(d_1 - r)$-by-r' random matrix. Therefore, $\|E_2\bar{V}\|_{\mathrm{sp}} \leq \sqrt{d_1}$ with probability at least $1 - C\exp(-cd_1)$. Therefore, combining all the bounds into (17), we proved that with probability at least $1 - C\exp(-cd_1)$,

$$\begin{aligned}\|Y_1\bar{V}\|_{\mathrm{sp}} &\leq \left(\frac{\sigma_{r'}(X)}{\sigma_{r'}(Y_1)} + 1\right)\left(\sqrt{d_1} + \|X_\perp\|_{\mathrm{sp}}\right) \\ &\leq 3(\sqrt{d_1} + \|X_\perp\|_{\mathrm{sp}}), \end{aligned} \tag{19}$$

where the first inequality uses the fact $\sigma_{r'}(X) \geq \sigma_{r'+1}(X)$ and the last inequality is from (14) and $\|X_{1,\perp}\|_{\mathrm{sp}} \vee \|X_{2,\perp}\|_{\mathrm{sp}} \leq \|X_\perp\|_{\mathrm{sp}}$.

Finally, plugging inequalities (15),(16), and (19) into (13) yeilds,

$$\|\hat{U}_{r\perp} U_{r'}\|_{\mathrm{sp}} \leq C \left(\frac{\sqrt{d_1} + \|X_\perp\|_{\mathrm{sp}}}{\sigma_{r'}(X)} + \frac{\sqrt{d_1 d_2} + \|X_\perp\|_{\mathrm{sp}}^2}{\sigma_{r'}^2(X)} \right).$$

$\square$

References

[1] T Tony Cai and Anru Zhang. Rate-optimal perturbation bounds for singular subspaces with applications to high-dimensional statistics. *The Annals of Statistics*, 46(1):60–89, 2018.

[2] Rungang Han, Yuetian Luo, Miaoyan Wang, and Anru R Zhang. Exact clustering in tensor block model: Statistical optimality and computational limit. *arXiv preprint arXiv:2012.09996*, 2020.