Supplementary for "On the Support Recovery of Jointly Sparse Gaussian Sources Via Sparse Bayesian Learning"

Saurabh Khanna, *Member, IEEE* and Chandra R. Murthy, *Senior Member, IEEE*

The contents of this supplementary document are as follows.

1) A complete, detailed proof of a modified version of Theorem 5 in *On the Restricted Isometry Property of Centered Self Khatri-Rao Products* [1] by Fengler et al. The modified theorem is stated as Theorem 1.
2) A new result characterizing the gap between the restricted minimum singular values of the centered, rescaled self Khatri-Rao product of a matrix and its uncentered self Khatri-Rao product variant. This result is stated as Theorem 4.

The notation used in this document is carried-forward from the parent paper *On the Support Recovery of Jointly Sparse Gaussian Sources Via Sparse Bayesian Learning* by S. Khanna et al.

Let $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n] \in \mathbb{R}^{m \times n}$ be an $m \times n$ matrix with $\mathbf{a}_i$ denoting the i^{th} column vector in $\mathbf{A}$. The centered self Khatri-Rao (KR) product of $\mathbf{A}$ is denoted by $\mathcal{A} \in \mathbb{R}^{m^2 \times n}$ whose i^{th} column is given by

$$\mathcal{A}_i = \kappa(m)\text{vec}(\mathbf{a}_i\mathbf{a}_i^T - \mathbf{I}_m), \tag{1}$$

where

$$\kappa(m) = \frac{m^2}{\mathbb{E}\left\{\|\text{vec}(\mathbf{a}_i\mathbf{a}_i^T - \mathbf{I}_m)\|_2^2\right\}} \tag{2}$$

serves as a normalization constant which ensures that the columns of $\mathcal{A}$ are normalized to unit ℓ_2-norm in expectation. In Theorem 1 below, we state a modified version of Theorem 5 in [1], which provides a probabilistic bound for the restricted isometry constant (RIC) of centered self KR product $\mathcal{A}$ as defined in (1) when $\mathbf{A}$ is comprised of i.i.d. sub-Gaussian entries.

Theorem 1. *Let m, n, k be positive integers such that $m^2 \leq n$ and $1 \leq k \leq m^2$. Let $\mathbf{A} = (A_{ij})$ be a random matrix with sub-Gaussian i.i.d. entries, such that $\mathbb{E}A_{ij} = 0$, $\mathbb{E}A_{ij}^2 = 1$ and $\|A_{ij}\|_{\psi_2} \leq B$. Let $\mathcal{A}$ be the centered and rescaled self KR product of $\mathbf{A}$ as defined in* (1)*. Then, for any $b > 0$, the RIP constant of order k of $\frac{\mathcal{A}}{m}$ satisfies*

$$\delta_k\left(\frac{\mathcal{A}}{m}\right) \leq \delta \tag{3}$$

for any $\delta > 0$ with probability larger than

$$1 - \max\left(\exp\left(-c\sqrt{n}\right), \frac{1}{(en)^{c\sqrt{k}/2}}\right) - \frac{2C}{n^b} \tag{4}$$

as long as

$$k \leq \frac{cm^2}{\log^2(e\frac{n}{cm^2})}, \tag{5}$$

and

$$m \geq \max\left(\frac{(b+2)\sqrt{2}B^2\log n}{c\sqrt{\delta}}, 1 + (\mathbb{E}a^4 - 1)(\frac{6}{\delta} - 1)\right), \tag{6}$$

where $c \leq \min\left((1/e)^{4/3}, \left(\frac{\delta}{6C\xi^2}\right)^2\right)$, $\xi = c'B^2 + 1$ and c', C are universal positive constants. Here, a is a generic random variable with the same distribution as the i.i.d. entries of $\mathbf{A}$.

Proof. See Appendix A. □

In Theorem 2, we restate an important result from [1] about the concentration of ℓ_2-norm of columns of the centered, self-KR product $\mathcal{A}$. Its immediately following corollary is useful in proving Theorem 1.

Theorem 2. *(Theorem 4 in [1]) Let $\mathbf{A} = (A_{ij})$ be a random matrix with sub-Gaussian iid entries, satisfying $\|A_{ij}\|_{\psi_2} \leq B$ and $\mathbb{E}A_{ij} = 0$ and normalized such that $\mathbb{E}A_{ij}^2 = 1$. Further, let $\{\mathcal{A}_i\}_{i=1}^n$ be the columns of the centered self KR product of $\mathbf{A}$. Then, it holds:*

$$\mathbb{P}\left(\max_{i\leq n}\left|\frac{\|\mathcal{A}_i\|_2^2}{m^2} - 1\right| \geq t\right) \leq C\exp\left(\log n - \frac{c}{B^2}\sqrt{t}m\right) \tag{7}$$

if m satisfies

$$m \geq 1 + (\mathbb{E}a^4 - 1)(3/t - 1), \tag{8}$$

with $a \sim P_a$, the distribution of the i.i.d. entries of $\mathbf{A}$. Here, C and c are universal positive constants.

Proof. See proof of Theorem 4 in [1]. □

In the following corollary of Theorem 2, we present two tail bounds that are ultimately used to prove Theorem 1.

Corollary 1. *For* $\mathbf{A}$ *and* $\mathcal{A}$ *as defined in Theorem 2, it follows that*

$$\mathbb{P}\left(\max_{i\le n}\left|\frac{\|\mathcal{A}_i\|_2^2}{m^2}-1\right|\ge\frac{\delta}{2}\right)\le\frac{C}{n^{b+1}}, \tag{9}$$

and

$$\mathbb{P}\left(\max_{i\le n}\|\mathcal{A}_i\|_2\ge\sqrt{1+\delta/2}m\right)\le\frac{C}{n^{b}}, \tag{10}$$

provided $m\ge\max\left(\frac{(b+2)\sqrt{2}B^2\log n}{c\sqrt{\delta}},1+(\mathbb{E}a^4-1)(\frac{6}{\delta}-1)\right)$*, with* $a\sim P_a$*, the distribution of the i.i.d. entries of* $\mathbf{A}$*, and* $c,C>0$ *being universal constants.*

Proof. By invoking Theorem 2 with $t=\frac{\delta}{2}$, it follows that

$$\mathbb{P}\left(\max_{i\le n}\left|\frac{\|\mathcal{A}_i\|_2^2}{m^2}-1\right|\ge\frac{\delta}{2}\right)\le C\exp\left(\log n-\frac{c\sqrt{\delta}m}{\sqrt{2}B^2}\right)\le C\exp\left(-(b+1)\log n\right)=\frac{C}{n^{b+1}}, \tag{11}$$

provided $\frac{c\sqrt{\delta}m}{\sqrt{2}B^2}\ge(b+2)\log n$, and $m\ge1+(\mathbb{E}a^4-1)(\frac{6}{\delta}-1)$. Or equivalently, the above tail bound in (11) holds true as long as $m\ge\max\left(\frac{(b+2)\sqrt{2}B^2\log n}{c\sqrt{\delta}},1+(\mathbb{E}a^4-1)(\frac{6}{\delta}-1)\right)$.

Next, the tail bound in (10) can be obtained by noting that

$$\mathbb{P}\left(\max_{i\le n}\|\mathcal{A}_i\|_2\ge\sqrt{1+\delta/2}m\right)\le n\mathbb{P}\left(\frac{\|\mathcal{A}_i\|_2^2}{m^2}-1\ge(1+\delta/2)-1\right) \tag{12}$$

$$=n\mathbb{P}\left(\frac{\|\mathcal{A}_i\|_2^2}{m^2}-1\ge\delta/2\right) \tag{13}$$

$$\le n\mathbb{P}\left(\left|\frac{\|\mathcal{A}_i\|_2^2}{m^2}-1\right|\ge\delta/2\right) \tag{14}$$

$$\le Cn\exp\left(\log n-\frac{c\sqrt{\delta}m}{\sqrt{2}B^2}\right) \tag{15}$$

$$\le Cn\exp\left(-(b+1)\log n\right)=\frac{C}{n^{b}}. \tag{16}$$

In the above, the first inequality is the union bound. The penultimate inequality follows from Theorem 2 with t set equal to $\delta/2$. The final inequality holds true due to the assumption that $m\ge\frac{\sqrt{2}(b+2)B^2\log n}{c\sqrt{\delta}}$. □

Finally, we restate Theorem 3.2 from [2] as Theorem 3 below. The theorem characterizes the restricted isometry property of randomly constructed matrices containing independent columns.

Theorem 3 (Theorem 3.2 in [2]). *Let $m \geq 1$ and s, N be integers such that $1 \leq s \leq \min(N, m)$. Let $X_1, X_2, \ldots, X_N \in \mathbb{R}^m$ be independent ψ_1 random vectors normalized such that $\mathbb{E}\left\{\|X_i\|^2\right\} = m$ and let $\psi = \max_{i \leq N} \|X_i\|_{\psi_1}$. Let $\theta' \in (0,1)$, $K, K' \geq 1$ and set $\xi = \psi K + K'$. Then, for matrix A with columns X_i, $A := (X_1 | \ldots | X_N)$*

$$\delta_s\left(\frac{A}{\sqrt{m}}\right) \leq C\xi^2\sqrt{\frac{s}{m}}\log\left(\frac{eN}{s\sqrt{\frac{s}{m}}}\right) + \theta'$$

holds with probability larger than

$$1 - \exp\left(-cK\sqrt{s}\log\left(\frac{eN}{s\sqrt{\frac{s}{m}}}\right)\right) - \mathbb{P}\left(\max_{i \leq N}\|X_i\|_2 \geq K'\sqrt{m}\right) - \mathbb{P}\left(\max_{i \leq N}\left|\frac{\|X_i\|_2^2}{m} - 1\right| \geq \theta'\right), \tag{17}$$

where $C, c > 0$ are universal constants.

Proof. See proof of Theorem 3.2 in [2]. □

We conclude our discussion on RIP of self Khatri-Rao product matrices by stating the relation between the restricted minimum singular values of the non-centered and the centered versions of the self Khatri-Rao product of Gaussian matrices containing i.i.d. entries.

Theorem 4. *Let $\mathbf{A} = (A_{ij})$ be a random matrix with i.i.d. $\mathcal{N}(0,1)$ entries. Let $\mathcal{A}$ be the centered and rescaled self KR product of $\mathbf{A}$ as defined in (1). Then, for any $\xi > 1$, we have*

$$\beta_k\left(\frac{\mathbf{A} \odot \mathbf{A}}{m}\right) \geq \beta_k\left(\frac{\mathcal{A}}{m}\right) - \frac{2\xi k \log n}{cm^2}, \tag{18}$$

with probability exceeding $1 - 2/n^{\xi - 1}$, provided $m \geq \frac{16\xi \log n}{c}$, where c is a universal numerical constant. Here, $\beta_k(\cdot)$ denotes the squared value of the k^{th} order restricted minimum singular value[1] of the input matrix

Proof. See Appendix C. □

[1]For an $m \times n$ matrix $\mathbf{X}$, and a positive integer $k \leq n$, the k^{th} order restricted minimum singular value of $\mathbf{X}$ is defined as the smallest among the singular values of all $m \times k$ submatrices obtained by sampling k or fewer columns of $\mathbf{X}$, i.e., $\min_{\mathcal{S} \subseteq [n], |\mathcal{S}| \leq k} \sqrt{\lambda_{\min}\left(\mathbf{X}_{\mathcal{S}}^H \mathbf{X}_{\mathcal{S}}\right)}$, where $\lambda_{\min}(\cdot)$ denotes the minimum eigenvalue of the input matrix.

According to Theorem 4, the gap between the restricted minimum singular values of the noncentered and centered self Khatri-Rao products can be made vanishingly small. A particularly interesting case is $m = \Omega(\xi\sqrt{k}\log n)$, wherein it is guaranteed that

$$\beta_k\left(\frac{\mathbf{A}\odot\mathbf{A}}{m}\right) \geq \beta_k\left(\frac{\mathcal{A}}{m}\right) - \Theta\left(\frac{1}{\xi}\right) \tag{19}$$

with probability exceeding $1 - 2/n^{\xi-1}$.

Appendix

A. Proof of Theorem 1

[1, Theorem 3] shows that the columns of $\mathcal{A}$ are subexponential such that $\|\mathcal{A}_i\|_{\psi_1} \leq c'B^2$. So the prerequisites of Theorem 3 are fulfilled with $\psi = c'B^2$, for some absolute constant $c' > 0$, and the variable m (in Theorem 3) replaced with m^2. We set $\theta' = \delta/2$ and $K' = \sqrt{1+\delta/2}$. Furthermore, we can set $K = 1$ such that $\xi = \psi K + K' = c'B^2 + 1$. Then, by invoking Theorem 3, it follows that

$$\delta_k\left(\frac{\mathcal{A}}{m}\right) \;\leq\; \underbrace{C\xi^2\sqrt{\frac{k}{m^2}}\log\left(\frac{en}{k\sqrt{\frac{k}{m^2}}}\right)}_{D} + \frac{\delta}{2} \tag{20}$$

with probability exceeding

$$1 - \exp\left(-cK\sqrt{k}\log\left(\frac{en}{k\sqrt{\frac{k}{m^2}}}\right)\right) - \mathbb{P}\left(\max_{i\leq n}\|\mathcal{A}_i\|_2 \geq \sqrt{1+\delta/2}\,m\right) - \mathbb{P}\left(\max_{i\leq n}\left|\frac{\|\mathcal{A}_i\|_2^2}{m^2} - 1\right| \geq \frac{\delta}{2}\right). \tag{21}$$

For $m \geq \max\left(\frac{(b+2)\sqrt{2}B^2\log n}{c\sqrt{\delta}}, 1 + (\mathbb{E}a^4 - 1)(\frac{6}{\delta} - 1)\right)$, by substituting the tail probabilities from Corollary 1 in (21), it follows that

$$\mathbb{P}\left(\delta_k\left(\frac{\mathcal{A}}{m}\right) \leq D + \frac{\delta}{2}\right) \geq 1 - \exp\left(-cK\sqrt{k}\log\left(\frac{en}{k\sqrt{\frac{k}{m^2}}}\right)\right) - \frac{2C}{n^b}. \tag{22}$$

Let $k \leq k^* \triangleq cm^2/\log^2(e\frac{n}{cm^2})$ for any $0 < c \leq 1$. Note that the conditions $c \leq 1$ and $n \geq cm^2$ guarantee that $\log(e\frac{n}{cm^2}) \geq 1$. Further, as argued in Appendix B, the second term D in (20) increases monotonically with k in the interval $[0, k^*]$, provided $c \leq (1/e)^{4/3}$ and $n \geq cm^2$.

Therefore for $k \leq k^*$, D attains its maximum value at $k = k^*$ for an appropriately chosen value for the constant c. Then, by plugging $k = k^*$ into (20) we see that the RIP constant satisfies

$$\begin{aligned}
\delta_k\left(\frac{\mathcal{A}}{m}\right) &\leq D|_{k=k^*} + \frac{\delta}{2} && (23)\\
&= C\xi^2\sqrt{c}\frac{\log\left(e\frac{n}{c^{3/2}m^2}\log^3(e\frac{n}{cm^2})\right)}{\log(e\frac{n}{cm^2})} + \frac{\delta}{2} && (24)\\
&\leq C\xi^2\sqrt{c}\frac{\log\left(e(\frac{n}{cm^2})^{3/2}\log^3(e\frac{n}{cm^2})\right)}{\log(e\frac{n}{cm^2})} + \frac{\delta}{2} && (25)\\
&= C\xi^2\sqrt{c}\left(\frac{3}{2} + \frac{3\log\log e\frac{n}{cm^2}}{\log e\frac{n}{cm^2}}\right) + \frac{\delta}{2} && (26)\\
&\leq C\xi^2\sqrt{c}\left(\frac{3}{2} + \frac{3}{e}\right) + \frac{\delta}{2} && (27)\\
&\leq 3C\xi^2\sqrt{c} + \frac{\delta}{2} \quad\leq\quad \delta, && (28)
\end{aligned}$$

where in the first inequality we made use of $m^2 \leq n$ and in the second to the last inequality we used $\log\log x/\log x \leq\ 1/e$. The final inequality holds for $c \leq \left(\frac{\delta}{6C\xi^2}\right)^2$. From (21), this bound fails with probability:

$$\begin{aligned}
\mathbb{P}\left(\delta_k\left(\frac{\mathcal{A}}{m}\right) \geq \delta\right) &\leq \exp\left(-c\sqrt{k}\log\left(e\frac{nm}{k^{3/2}}\right)\right) + \frac{2C}{n^b} && (29)\\
&\leq \underbrace{\exp\left(-c\sqrt{k}\log\left(e\frac{n}{k}\right)\right)}_{Z} + \frac{2C}{n^b}, && (30)
\end{aligned}$$

where in the second inequality it was used that $k \leq m^2$.

For $k \geq \sqrt{n}$, we note that the first term in (30) can be bounded as

$$Z \leq \exp\left(-c\sqrt{n}\log\left(e\frac{n}{k}\right)\right) \leq \exp\left(-c\sqrt{n}\right). \tag{31}$$

On the other hand, for $k < \sqrt{n}$, we have

$$Z \leq \exp\left(-c\sqrt{k}\log\left(e\sqrt{n}\right)\right) \leq \frac{1}{(en)^{c\sqrt{k}/2}}. \tag{32}$$

Combining (30), (31) and (32), we have

$$\mathbb{P}\left(\delta_k\left(\frac{\mathcal{A}}{m}\right) \geq \delta\right) \leq \max\left(\exp\left(-c\sqrt{n}\right), \frac{1}{(en)^{c\sqrt{k}/2}}\right) + \frac{2C}{n^b}, \tag{33}$$

provided $k \leq \frac{cm^2}{\log^2(e\frac{n}{cm^2})}$, and $\sqrt{n} \geq m \geq \max\left(\frac{(b+1)\sqrt{2}B^2\log n}{c\sqrt{\delta}}, 1 + (\mathbb{E}a^4 - 1)(\frac{6}{\delta} - 1)\right)$ with $c \leq \min\left((1/e)^{4/3}, \left(\frac{\delta}{6C\xi^2}\right)^2\right)$.

B. Characterization of D in (30) *as a function of k*

In this appendix, we independently show that D in (30) is a monotonic increasing function with respect to k for $k \leq k^* \triangleq cm^2/\log^2(e\frac{n}{cm^2})$, provided $c \leq (1/e)^{4/3}$ and $n \geq m^2$. Note that, D can be expressed as

$$D = \frac{C\xi^2}{m} g(t)\big|_{t=\sqrt{k}}$$

$$\text{where } g(t) = t \log\left(\frac{enm}{t^3}\right). \tag{34}$$

The first and second derivatives of $g(t)$ with respect to t, denoted $g'(t)$ and $g''(t)$, are

$$g'(t) = \log(enm) - 3\log t - 3 \text{ and } g''(t) = -\frac{3}{t}. \tag{35}$$

Since $g'(0) = +\infty$ and $g''(t)$ is negative for $t > 0$, $g(t)$ is monotonically increasing and strictly concave for $0 < t \leq t^*$ as long as $g'(t^*) \geq 0$. We now argue that $g'(t^*) \geq 0$ for $t^* = \sqrt{k^*} := \frac{\sqrt{c}m}{\log\left(e\frac{n}{cm^2}\right)}$, in turn implying that $g(t)$ increases monontonically with t in the interval $(0, t^*]$. We note that

$$\begin{aligned} g'(t^*)\big|_{t^*=\sqrt{k^*}} &= \left(\log(enm) - 3\log t^* - 3\right)\big|_{t^*=\sqrt{k^*}} \\ &= \log(enm) - 3\log\left(\frac{\sqrt{c}m}{\log\left(\frac{en}{cm^2}\right)}\right) - 3 \\ &\geq \log\left(\frac{nm}{e^2}\right) - 3\log\left(\sqrt{c}m\right) \qquad \left(\text{as } \log\left(\frac{en}{cm^2}\right) \geq 1\right) \end{aligned} \tag{36}$$

is nonnegative if $\frac{nm}{e^2} > (cm^2)^{3/2}$, or equivalently, if $c \leq \left(\frac{n}{m^2e^2}\right)^{2/3}$. Since, in Corollary 1, we assume that $m^2 \leq n$, it follows that choosing $c \leq (1/e)^{\frac{4}{3}}$ guarantees $g'(\sqrt{k^*}) \geq 0$, and consequently $g(t)$ is monotonically increasing in the interval $0 \leq t \leq \sqrt{k^*} := \frac{\sqrt{c}m}{\log\left(e\frac{n}{cm^2}\right)}$. Therefore, for $t \leq \sqrt{k^*}$, or equivalently, $k \leq k^*$, we have

$$\begin{aligned} D = \frac{C\xi^2}{m} g(t) &\leq \frac{C\xi^2}{m} g(\sqrt{k^*}) \\ &= C\xi^2\sqrt{c}\frac{\log\left(\frac{en}{c^{3/2}m^2}\log^3\left(e\frac{n}{cm^2}\right)\right)}{\log(e\frac{n}{cm^2})} \\ &\leq C\xi^2\sqrt{c}\frac{\log\left(e(\frac{n}{cm^2})^{3/2}\log^3(e\frac{n}{cm^2})\right)}{\log(e\frac{n}{cm^2})} \qquad (\text{as } \frac{n}{m^2} \geq 1). \end{aligned} \tag{37}$$

C. Proof of Theorem 4

(Restricted singular values of non-centered self Khatri-Rao product of Gaussian matrices)

Proof. Let $\mathcal{A}$ denote the centered self Khatri-Rao product of $\mathbf{A}$, whose columns are defined according to (1). Then, one can write

$$\mathcal{A} = \kappa(m)\left(\mathbf{A} \odot \mathbf{A} - \mathbf{B}\right), \tag{38}$$

where $\mathbf{B} = \text{vec}(\mathbf{I}_m)\mathbf{1}_n^T$. Let $\mathcal{S} \subset [n]$ denote an arbitrary k-sized index set with distinct elements. Then, from (38), it follows that

$$\mathcal{A}_{\mathcal{S}} = \kappa(m)\left(\mathbf{A}_{\mathcal{S}} \odot \mathbf{A}_{\mathcal{S}} - \mathbf{B}_{\mathcal{S}}\right), \tag{39}$$

and furthermore,

$$\begin{aligned}
&(\mathbf{A}_{\mathcal{S}} \odot \mathbf{A}_{\mathcal{S}})^T(\mathbf{A}_{\mathcal{S}} \odot \mathbf{A}_{\mathcal{S}}) && (40)\\
&= \frac{\mathcal{A}_{\mathcal{S}}^T\mathcal{A}_{\mathcal{S}}}{\kappa(m)^2} - \mathbf{B}_{\mathcal{S}}^T\mathbf{B}_{\mathcal{S}} + \mathbf{B}_{\mathcal{S}}^T(\mathbf{A}_{\mathcal{S}} \odot \mathbf{A}_{\mathcal{S}}) + (\mathbf{A}_{\mathcal{S}} \odot \mathbf{A}_{\mathcal{S}})^T\mathbf{B}_{\mathcal{S}} && (41)\\
&= \frac{\mathcal{A}_{\mathcal{S}}^T\mathcal{A}_{\mathcal{S}}}{\kappa(m)^2} - m\mathbf{1}_k\mathbf{1}_k^T + \mathbf{B}_{\mathcal{S}}^T(\mathbf{A}_{\mathcal{S}} \odot \mathbf{A}_{\mathcal{S}}) + (\mathbf{A}_{\mathcal{S}} \odot \mathbf{A}_{\mathcal{S}})^T\mathbf{B}_{\mathcal{S}} && (42)\\
&= \frac{\mathcal{A}_{\mathcal{S}}^T\mathcal{A}_{\mathcal{S}}}{\kappa(m)^2} - m\mathbf{1}_k\mathbf{1}_k^T + \mathbf{1}_k\mathbf{d}_{\mathcal{S}}^T + \mathbf{d}_{\mathcal{S}}\mathbf{1}_k^T && (43)\\
&= \frac{\mathcal{A}_{\mathcal{S}}^T\mathcal{A}_{\mathcal{S}}}{\kappa(m)^2} + \left(\mathbf{1}_k\mathbf{d}_{\mathcal{S}}^T - \frac{m}{2}\mathbf{1}_k\mathbf{1}_k^T\right) + \left(\mathbf{1}_k\mathbf{d}_{\mathcal{S}}^T - \frac{m}{2}\mathbf{1}_k\mathbf{1}_k^T\right)^T && (44)\\
&= \frac{\mathcal{A}_{\mathcal{S}}^T\mathcal{A}_{\mathcal{S}}}{\kappa(m)^2} + \mathbf{1}_k\left(\mathbf{d}_{\mathcal{S}} - \frac{m}{2}\mathbf{1}_k\right)^T + \left(\mathbf{d}_{\mathcal{S}} - \frac{m}{2}\mathbf{1}_k\right)\mathbf{1}_k^T, && (45)
\end{aligned}$$

where $\mathbf{d} \triangleq [\|\mathbf{a}_1\|_2^2\ \|\mathbf{a}_2\|_2^2\ \ldots\ \|\mathbf{a}_n\|_2^2]^T \in \mathbb{R}_+^n$ is defined to be the vector containing the squared ℓ_2-norms of the columns of $\mathbf{A}$, and $\mathbf{d}_{\mathcal{S}}$ denotes the k-dimensional subvector of $\mathbf{d}$ containing entries indexed by the support set $\mathcal{S}$.

Then, by the Weyl inequality [3], it follows from (45) that

$$\begin{aligned}
\lambda_{\min}\left(\frac{(\mathbf{A}_{\mathcal{S}} \odot \mathbf{A}_{\mathcal{S}})^T(\mathbf{A}_{\mathcal{S}} \odot \mathbf{A}_{\mathcal{S}})}{m^2}\right) \;\geq\; & \frac{1}{\kappa(m)^2}\lambda_{\min}\left(\frac{\mathcal{A}_{\mathcal{S}}^T\mathcal{A}_{\mathcal{S}}}{m^2}\right)\\
+\; & \frac{1}{m^2}\lambda_{\min}\left(\mathbf{1}_k\left(\mathbf{d}_{\mathcal{S}} - \frac{m}{2}\mathbf{1}_k\right)^T + \left(\mathbf{d}_{\mathcal{S}} - \frac{m}{2}\mathbf{1}_k\right)\mathbf{1}_k^T\right),
\end{aligned} \tag{46}$$

where $\lambda_{\min}(\cdot)$ denotes the minimum eigenvalue of the input matrix. By using Lemma 1, the minimum eigenvalue of $\mathbf{1}_k\left(\mathbf{d}_{\mathcal{S}} - \frac{m}{2}\mathbf{1}_k\right)^T + \left(\mathbf{d}_{\mathcal{S}} - \frac{m}{2}\mathbf{1}_k\right)\mathbf{1}_k^T$ can be bounded as shown below.

$$\lambda_{\min}\left(\mathbf{1}_k\left(\mathbf{d}_{\mathcal{S}} - \frac{m}{2}\mathbf{1}_k\right)^T + \left(\mathbf{d}_{\mathcal{S}} - \frac{m}{2}\mathbf{1}_k\right)\mathbf{1}_k^T\right) \;\geq\; \mathbf{v}^T\mathbf{w} - \|\mathbf{v}\|_2\|\mathbf{w}\|_2, \tag{47}$$

where $\mathbf{v} \triangleq \mathbf{1}_k$ and $\mathbf{w} \triangleq \mathbf{d}_{\mathcal{S}} - \frac{m}{2}\mathbf{1}_k$. Since $\mathbf{v}$ is positive and if $\mathbf{w}$ is also positive, then by invoking the Cassel's inequality [4], the lower bound in (47) can be relaxed further as

$$\lambda_{\min}\left(\mathbf{1}_k\left(\mathbf{d}_{\mathcal{S}} - \frac{m}{2}\mathbf{1}_k\right)^T + \left(\mathbf{d}_{\mathcal{S}} - \frac{m}{2}\mathbf{1}_k\right)\mathbf{1}_k^T\right) \;\; \geq \;\; \mathbf{v}^T\mathbf{w} - \frac{h+H}{2\sqrt{hH}}\mathbf{v}^T\mathbf{w}, \tag{48}$$

where $h \leq \frac{\mathbf{v}_i}{\mathbf{w}_i} \leq H$.

We now show that $\mathbf{w}$ is a strictly positive vector with high probability, and with each of its entries concentrating around $m/2$. Since $\mathbf{A}$ contains i.i.d. $\mathcal{N}(0,1)$ entries, by invoking the Hanson-Wright concentration inequality [5], and taking the union bound over all columns of $\mathbf{A}$, we have

$$\begin{aligned}\mathbb{P}\left(\max_{i\in[n]}\left|\|\mathbf{a}_i\|_2^2 - m\right| \geq \alpha\right) &= \mathbb{P}\left(\bigcup_{i\in[n]}\left\{\left|\|\mathbf{a}_i\|_2^2 - m\right| \geq \alpha\right\}\right)\\ &\leq n\mathbb{P}\left(\left|\mathbf{a}_1^T\mathbf{I}_m\mathbf{a}_1 - m\right| \geq \alpha\right)\\ &\leq 2ne^{-c\alpha^2/m},\end{aligned} \tag{49}$$

where c is a numerical constant. By substituting $\alpha = \sqrt{\frac{\xi m \log n}{c}}$ in (49), we obtain

$$\mathbb{P}\left(\max_{i\in[n]}\left|\|\mathbf{a}_i\|_2^2 - m\right| \geq \sqrt{\frac{\xi m \log n}{c}}\right) \leq \frac{2}{n^{\xi-1}}. \tag{50}$$

From (50), we can conclude that

$$\left|\|\mathbf{a}_i\|_2^2 - m\right| < \sqrt{\frac{\xi m \log n}{c}} \qquad \forall i \in [n] \tag{51}$$

with probability exceeding $1 - \frac{2}{n^{\xi-1}}$. Consequently, for all k-sparse supports $\mathcal{S}$, the vector $\mathbf{w} = \mathbf{d}_{\mathcal{S}} - \frac{m}{2}\mathbf{1}_k$ contains entries satisfying

$$\frac{m}{2} - \Delta \;<\; \mathbf{w}_i = \|\mathbf{a}_{\mathcal{S}(i)}\|_2^2 - \frac{m}{2} \;<\; \frac{m}{2} + \Delta, \qquad \text{for } 1 \leq i \leq k, \tag{52}$$

with probability exceeding $1 - \frac{2}{n^{\xi-1}}$ and $\Delta \triangleq \sqrt{\frac{\xi m \log n}{c}}$. From (52), it is evident that for $m \geq \frac{4\xi \log n}{c}$, $\mathbf{w}$ is a strictly positive vector with high probability, for all k-sparse supports $\mathcal{S}$.

Noting that $\mathbf{v} = \mathbf{1}_k$ in (53), and by using (52), one can choose $h = 1/(\frac{m}{2} + \Delta)$ and $H = 1/(\frac{m}{2} - \Delta)$, to obtain

$$\lambda_{\min}\left(\mathbf{1}_k\left(\mathbf{d}_{\mathcal{S}} - \frac{m}{2}\mathbf{1}_k\right)^T + \left(\mathbf{d}_{\mathcal{S}} - \frac{m}{2}\mathbf{1}_k\right)\mathbf{1}_k^T\right) \;\; \geq \;\; -(\mathbf{1}_k^T\mathbf{w})\left(\frac{\left(\frac{m}{2}+\Delta\right) + \left(\frac{m}{2}-\Delta\right)}{2\sqrt{\left(\frac{m}{2}+\Delta\right)\left(\frac{m}{2}-\Delta\right)}} - 1\right)$$

$$= -(\mathbf{1}_k^T \mathbf{w}) \left(\frac{1}{\sqrt{1 - \left(\frac{2\Delta}{m}\right)^2}} - 1 \right)$$
$$\geq -k \left(\frac{m}{2} + \Delta\right) \left(\frac{1}{\sqrt{1 - \left(\frac{2\Delta}{m}\right)^2}} - 1 \right). \quad (53)$$

Using the binomial expansion $\frac{1}{\sqrt{1-\left(\frac{2\Delta}{m}\right)^2}} = 1 + \frac{1}{2}\left(\frac{2\Delta}{m}\right)^2 + \frac{3}{8}\left(\frac{2\Delta}{m}\right)^4 + \frac{5}{16}\left(\frac{2\Delta}{m}\right)^6 + \ldots$, the lower bound in (53) simplifies as

$$\begin{aligned}
&\lambda_{\min}\left(\mathbf{1}_k \left(\mathbf{d}_{\mathcal{S}} - \frac{m}{2}\mathbf{1}_k\right)^T + \left(\mathbf{d}_{\mathcal{S}} - \frac{m}{2}\mathbf{1}_k\right)\mathbf{1}_k^T\right) \\
&\qquad \geq -k\left(\frac{m}{2} + \Delta\right)\left(\frac{1}{2}\left(\frac{2\Delta}{m}\right)^2 + \frac{3}{8}\left(\frac{2\Delta}{m}\right)^4 + \frac{5}{16}\left(\frac{2\Delta}{m}\right)^6 + \ldots\right) \\
&\qquad \geq -k\left(\frac{m}{2} + \Delta\right)\left(\frac{1}{2}\left(\frac{2\Delta}{m}\right)^2 + \frac{1}{2}\left(\frac{2\Delta}{m}\right)^4 + \frac{1}{2}\left(\frac{2\Delta}{m}\right)^6 + \ldots\right) \\
&\qquad = -k\left(\frac{m}{2} + \Delta\right)\left(\frac{2\Delta^2}{m^2}\right)\left(\frac{1}{1 - \left(\frac{2\Delta}{m}\right)^2}\right) \qquad (54)
\end{aligned}$$

for all k-sparse supports $\mathcal{S}$ with probability exceeding $1 - 2/n^{\xi-1}$. For $m \geq \frac{16\xi \log n}{c}$, we have $\frac{2\Delta}{m} < \frac{1}{2}$, and from (54), it follows that

$$\begin{aligned}
\lambda_{\min}\left(\mathbf{1}_k \left(\mathbf{d}_{\mathcal{S}} - \frac{m}{2}\mathbf{1}_k\right)^T + \left(\mathbf{d}_{\mathcal{S}} - \frac{m}{2}\mathbf{1}_k\right)\mathbf{1}_k^T\right) &\geq -k\left(\frac{m}{2} + \Delta\right)\left(\frac{2\Delta^2}{m^2}\right)\left(\frac{4}{3}\right) \\
&= -k\left(\frac{m}{2} + \sqrt{\frac{\xi m \log n}{c}}\right)\left(\frac{8\xi \log n}{3cm}\right) \\
&\geq -k\left(\frac{m}{2} + \frac{m}{4}\right)\left(\frac{8\xi \log n}{3cm}\right) \\
&= -\frac{2\xi k \log n}{c} \qquad (55)
\end{aligned}$$

for all k-sparse supports $\mathcal{S}$ with probability exceeding $1 - 2/n^{\xi-1}$.

Combining (46) and (55), and noting that $\kappa(m) = \frac{m}{m+1}$ for $A_{ij} \stackrel{i.i.d.}{\sim} \mathcal{N}(0,1)$, we have

$$\begin{aligned}
&\min_{\mathcal{S}\subseteq[n],|\mathcal{S}|\leq k} \lambda_{\min}\left(\frac{(\mathbf{A}_{\mathcal{S}} \odot \mathbf{A}_{\mathcal{S}})^T(\mathbf{A}_{\mathcal{S}} \odot \mathbf{A}_{\mathcal{S}})}{m^2}\right) \\
&\qquad \geq \min_{\mathcal{S}\subseteq[n],|\mathcal{S}|\leq k}\left(\frac{1}{\kappa(m)^2}\lambda_{\min}\left(\frac{\mathcal{A}_{\mathcal{S}}^T\mathcal{A}_{\mathcal{S}}}{m^2}\right)\right. \\
&\qquad\qquad \left. + \frac{1}{m^2}\lambda_{\min}\left(\mathbf{1}_k \left(\mathbf{d}_{\mathcal{S}} - \frac{m}{2}\mathbf{1}_k\right)^T + \left(\mathbf{d}_{\mathcal{S}} - \frac{m}{2}\mathbf{1}_k\right)\mathbf{1}_k^T\right)\right)
\end{aligned}$$

$$\geq \left(1+\frac{1}{m}\right)^2 \min_{\mathcal{S}\subseteq[n],|\mathcal{S}|\leq k} \lambda_{\min}\left(\frac{\mathcal{A}_{\mathcal{S}}^T\mathcal{A}_{\mathcal{S}}}{m^2}\right) + \frac{1}{m^2}\min_{\mathcal{S}\subseteq[n],|\mathcal{S}|\leq k} \lambda_{\min}\left(\mathbf{1}_k\left(\mathbf{d}_{\mathcal{S}}-\frac{m}{2}\mathbf{1}_k\right)^T + \left(\mathbf{d}_{\mathcal{S}}-\frac{m}{2}\mathbf{1}_k\right)\mathbf{1}_k^T\right)$$

$$\geq \min_{\mathcal{S}\subseteq[n],|\mathcal{S}|\leq k} \lambda_{\min}\left(\frac{\mathcal{A}_{\mathcal{S}}^T\mathcal{A}_{\mathcal{S}}}{m^2}\right) - \frac{1}{m^2}\left(\frac{2\xi k\log n}{c}\right) \tag{56}$$

with probability exceeding $1-\frac{2}{n^{\xi-1}}$, provided $m \geq \frac{16\xi\log n}{c}$.

□

Lemma 1. *Let* $\mathbf{U} = \mathbf{v}\mathbf{w}^T + \mathbf{w}\mathbf{v}^T$ *be symmetric matrix such that* $\mathbf{v}$ *and* $\mathbf{w}$ *are linearly independent. Then,* $\mathbf{U}$ *has rank exactly equal to* 2 *and its two nonzero eigenvalues* λ_1 *and* λ_2 *are*

$$\lambda_1 = \mathbf{v}^T\mathbf{w} + \|\mathbf{v}\|_2\|\mathbf{w}\|_2$$

$$\text{and } \lambda_2 = \mathbf{v}^T\mathbf{w} - \|\mathbf{v}\|_2\|\mathbf{w}\|_2.$$

Proof. Since $\mathbf{U}$ is sum of two rank one matrices, its rank is at most two, and it has at most two nonzero eigenvalues, say λ_1 and λ_2. Then, we have

$$\lambda_1 + \lambda_2 = \text{trace}(\mathbf{U}) = 2\mathbf{v}^T\mathbf{w} \tag{57}$$

$$\text{and} \quad \lambda_1^2 + \lambda_2^2 = \text{trace}(\mathbf{U}^2) = 2\|\mathbf{v}\|_2^2\|\mathbf{w}\|_2^2 + 2(\mathbf{v}^T\mathbf{w})^2. \tag{58}$$

Noting that $\lambda_1\lambda_2 = \frac{1}{2}\left((\lambda_1+\lambda_2)^2 - (\lambda_1^2+\lambda_2^2)\right)$, and using (57) and (58), we have

$$\lambda_1\lambda_2 = (\mathbf{v}^T\mathbf{w})^2 - \|\mathbf{v}\|_2^2\|\mathbf{w}\|_2^2. \tag{59}$$

From (59), it is evident from the Cauchy-Schwarz inequality that since $\mathbf{v}$ and $\mathbf{w}$ are linearly independent, $\lambda_1\lambda_2 < 0$, and therefore $\mathbf{U}$ has exactly two nonzero eigenvalues, which are of opposite signs.

Further, from (57) and (59) together, we can conclude that λ_1 and λ_2 are the two roots of the quadratic polynomial

$$f(x) = x^2 - 2(\mathbf{v}^T\mathbf{w})x + \left((\mathbf{v}^T\mathbf{w})^2 - \|\mathbf{v}\|_2^2\|\mathbf{w}\|_2^2\right), \tag{60}$$

and can be explicitly evaluated as

$$\lambda_1 = \frac{2\mathbf{v}^T\mathbf{w} + \sqrt{4(\mathbf{v}^T\mathbf{w})^2 - 4\left((\mathbf{v}^T\mathbf{w})^2 - \|\mathbf{v}\|_2^2\|\mathbf{w}\|_2^2\right)}}{2} = \mathbf{v}^T\mathbf{w} + \|\mathbf{v}\|_2\|\mathbf{w}\|_2$$

$$\text{and } \lambda_2 = \frac{2\mathbf{v}^T\mathbf{w} - \sqrt{4(\mathbf{v}^T\mathbf{w})^2 - 4\left((\mathbf{v}^T\mathbf{w})^2 - \|\mathbf{v}\|_2^2\|\mathbf{w}\|_2^2\right)}}{2} = \mathbf{v}^T\mathbf{w} - \|\mathbf{v}\|_2\|\mathbf{w}\|_2.$$

□

REFERENCES

[1] A. Fengler and P. Jung, "On the Restricted Isometry Property of centered self Khatri-Rao products," *CoRR*, vol. abs/1905.09245, 2019. [Online]. Available: http://arxiv.org/abs/1905.09245

[2] R. Adamczak, A. E. Litvak, A. Pajor, and N. Tomczak-Jaegermann, "Restricted Isometry Property of matrices with independent columns and neighborly polytopes by random sampling," *Constructive Approximation*, vol. 34, pp. 61–88, 2009.

[3] D. S. Bernstein, *Matrix Mathematics: Theory, Facts, and Formulas (Second Edition)*. Princeton University Press, 2009.

[4] S. S. Dragomir, "A counterpart of Schwarz's inequality in inner product spaces," *Mathematics eJournal*, 2003.

[5] M. Rudelson and R. Vershynin, "Hanson-Wright inequality and sub-gaussian concentration," *Electron. Commun. Probab.*, vol. 18, p. 9 pp., 2013.