Proof for the theorem in "Subspace principal angle preserving property of gaussian random projection", submitted to IEEE Data Science Workshop

Yuchen Jiao Xinyue Shen Gen Li Yuantao Gu*

This manuscript is for supporting our submission to IEEE Data Science Workshop, "Subspace principal angle preserving property of gaussian random projection" [1].

To make the proof easier to understand, we denote the kth principal angle between original subspaces $\mathcal{X}_1$ and $\mathcal{X}_2$ as θ_k, between projected subspaces $\mathcal{Y}_1$ and $\mathcal{Y}_2$ as ψ_k. Define

$$\overline{\sin^2\psi_k} = \left(1 - \frac{d_2}{n}\right)\sin^2\theta_k = 1 - \bar{\lambda}^2_{\mathcal{Y}_k} \tag{1}$$

as the estimate of $\sin^2\psi_k$. Then by replacing $\lambda^2_{\mathcal{Y}_k}$ in the Theorem 1 in [1] with $1-\sin^2\psi_k$, we can rewrite that theorem as below.

Theorem 1. Suppose $\mathcal{X}_1$, $\mathcal{X}_2 \subset \mathbb{R}^N$ are two subspaces with dimension $d_1 \le d_2$, respectively. Principal angles between them are denoted by $\theta_1, \cdots, \theta_{d_1}$. After Gaussian random projection: $\mathcal{X}_l \xrightarrow{\mathbf{\Phi}} \mathcal{Y}_l, l = 1, 2$, which is defined in Notation 1 in [1], we can get $\mathcal{Y}_1, \mathcal{Y}_2 \subset \mathbb{R}^n$, and principal angles between them are denoted by $\psi_1, \cdots, \psi_{d_1}$. $\overline{\sin^2\psi_k}$ is defined in (1). Then there exist positive constants c_1, c_2 depending only on ε, such that for any $n \ge c_1 d_2$,

$$\left|\sin^2\psi_k - \overline{\sin^2\psi_k}\right| \le \varepsilon \sin^2\theta_k \tag{2}$$

holds with probability at least $1 - \mathrm{e}^{-c_2 n}$.

Before proof, we define standard Gaussian random matrix as below.

Definition 1. (Standard Gaussian Random Matrix) A standard Gaussian random matrix $\mathbf{A} \in \mathbb{R}^{n\times k}$ has i.i.d. zero-mean Gaussian random entries with variance $1/n$. A normalized standard Gaussian random matrix is the column-normalized result of a standard Gaussian random matrix.

*The authors are with the Department of Electronic Engineering, Tsinghua University, Beijing 100084, CHINA. The corresponding author of this work is Y. Gu (Email: gyt@tsinghua.edu.cn).

1 Notation

Vectors and matrices are denoted by lower-case and upper-case letter, respectively, both in boldface. $\mathbf{V}^{\mathrm{T}}$ denotes matrix transposition. $\|\mathbf{v}\|$ denotes the l_2 norm of vector $\mathbf{v}$. $\sigma_k(\mathbf{V})$ denotes the kth singular value of matrix $\mathbf{V}$ in descending order, that is $\sigma_1(\mathbf{V}) \geq \sigma_2(\mathbf{V}) \geq \cdots$. Subspaces are denoted by $\mathcal{X}, \mathcal{Y}$ and $\mathcal{S}$. $\mathcal{S}^{\perp}$ denotes the orthogonal complement space of $\mathcal{S}$. $\mathcal{C}(\mathbf{V})$ denotes the column space of $\mathbf{V}$. $\mathbf{P}_{\mathcal{S}}(\mathbf{v})$ denotes the projection of vector $\mathbf{v}$ onto subspace $\mathcal{S}$. $\mathbf{S}^k$ denotes the unit sphere in a k-dimensional subspace. $\mathbf{U}_l$, $\mathbf{A}_l$, $\bar{\mathbf{A}}_{1,1:k}$, $\mathbf{a}_{l,k}$, $\bar{\mathbf{a}}_{1,k}$, $\mathbf{a}_{1,k}^{\perp}$, $l = 1, 2$, $k = 1, \cdots, d_1$ are defined in Notation 2 in [1]. For the ease of understanding, we define $\bar{\mathbf{a}}_{1,k}^{\perp}$ as the unit vector along the projection of $\mathbf{a}_{1,k}$ onto $\mathcal{Y}_2^{\perp}$ instead of $\mathbf{a}_{1,k}^{\perp}$ defined in Notation 2 in [1]. Also define $\bar{\mathbf{A}}_{1,1:k}^{\perp} := \left[\bar{\mathbf{a}}_{1,1}^{\perp}, \cdots, \bar{\mathbf{a}}_{1,k}^{\perp}\right]$ instead of $\mathbf{A}_{1,1:k}^{\perp}$ defined in Notation 2 in [1]. Define $\bar{\mathbf{A}}_{1,k:d_1} := \left[\bar{\mathbf{a}}_{1,k}, \cdots, \bar{\mathbf{a}}_{1,d_1}\right]$, $\bar{\mathbf{A}}_{1,k:d_1}^{\perp} := \left[\bar{\mathbf{a}}_{1,k}^{\perp}, \cdots, \bar{\mathbf{a}}_{1,d_1}^{\perp}\right]$.

2 Main body of the proof of Theorem 1

We start from calculating the sine of the kth principal angle between projected subspaces via singular value. Denote $\mathbf{V}_l$ as the orthonormal basis of $\mathcal{Y}_l$, $l = 1, 2$. Then according to Lemma 1 and Lemma 2 in [1], we have

$$\begin{aligned} \sin^2 \psi_k &= 1 - \sigma_k^2\left(\mathbf{V}_2^{\mathrm{T}}\mathbf{V}_1\right) = 1 - \max_{\mathbf{S}^k \subset \mathbb{R}^{d_1}} \min_{\mathbf{x} \in \mathbf{S}^k} \left\|\mathbf{V}_2^{\mathrm{T}}\mathbf{V}_1\mathbf{x}\right\|^2 \\ &= 1 - \max_{\mathbf{Q} \in \mathbb{R}^{d_1 \times k}} \min_{\substack{\|\mathbf{y}\|=1 \\ \mathbf{y} \in \mathcal{C}(\mathbf{V}_1\mathbf{Q})}} \left\|\mathbf{V}_2^{\mathrm{T}}\mathbf{y}\right\|^2, \end{aligned} \tag{3}$$

The last equation could be obtained by letting $\mathbf{y} := \mathbf{V}_1\mathbf{x}$, and $\mathbf{Q}$ is a full rank matrix. Noticing that $\left\|\mathbf{V}_2^{\mathrm{T}}\mathbf{y}\right\|$ is the norm of the projection of $\mathbf{y}$ onto $\mathcal{Y}_2$, we project $\mathbf{y}$ onto $\mathcal{Y}_2$ and its orthogonal complement space $\mathcal{Y}_2^{\perp}$ to get $\mathbf{P}_{\mathcal{Y}_2}(\mathbf{y})$ and $\mathbf{P}_{\mathcal{Y}_2^{\perp}}(\mathbf{y})$, respectively, as follows

$$\mathbf{y} = \mathbf{P}_{\mathcal{Y}_2}(\mathbf{y}) + \mathbf{P}_{\mathcal{Y}_2^{\perp}}(\mathbf{y}). \tag{4}$$

Considering that the norm of $\mathbf{y}$ equals 1, we have

$$\begin{aligned} \sin^2 \psi_k &= 1 - \max_{\mathbf{Q} \in \mathbb{R}^{d_1 \times k}} \min_{\substack{\|\mathbf{y}\|=1 \\ \mathbf{y} \in \mathcal{C}(\mathbf{V}_1\mathbf{Q})}} \left\|\mathbf{P}_{\mathcal{Y}_2}(\mathbf{y})\right\|^2 \\ &= \min_{\mathbf{Q} \in \mathbb{R}^{d_1 \times k}} \max_{\substack{\|\mathbf{y}\|=1 \\ \mathbf{y} \in \mathcal{C}(\mathbf{V}_1\mathbf{Q})}} \left\|\mathbf{P}_{\mathcal{Y}_2^{\perp}}(\mathbf{y})\right\|^2 \qquad (5) \\ &\leq \max_{\substack{\|\mathbf{y}\|=1 \\ \mathbf{y} \in \mathcal{C}(\mathbf{V}_1\mathbf{Q}_u)}} \left\|\mathbf{P}_{\mathcal{Y}_2^{\perp}}(\mathbf{y})\right\|^2, \qquad (6) \end{aligned}$$

where $\mathbf{Q}_u := \left[\mathbf{V}_1^{\mathrm{T}}\mathbf{a}_{1,1}, \cdots, \mathbf{V}_1^{\mathrm{T}}\mathbf{a}_{1,k}\right]$. Then the subspace $\mathcal{C}(\mathbf{V}_1\mathbf{Q}_u)$ is

$$\mathcal{Y}_{1,1:k} := \mathrm{span}\{\mathbf{a}_{1,1}, \cdots, \mathbf{a}_{1,k}\}. \tag{7}$$

By plugging (7) into (6), the upper bound of $\sin^2\psi_k$ is formulated into the following maximum problem.

$$\sin^2\psi_k \le \max_{\substack{\|\mathbf{y}\|=1\\ \mathbf{y}\in\mathcal{Y}_{1,1:k}}} \left\|\mathbf{P}_{\mathcal{Y}_2^\perp}(\mathbf{y})\right\|^2. \tag{8}$$

In order to get the lower bound of $\sin^2\psi_k$, using (4) in [1] and following the same approach, we could derive step by step the counterparts of (3), (5), (6) and (8) as

$$\begin{aligned}\sin^2\psi_k &= 1-\sigma_k^2\left(\mathbf{V}_2^{\mathrm{T}}\mathbf{V}_1\right) = 1-\min_{\mathbf{S}^{d_1-k+1}\subset\mathbb{R}^{d_1}}\max_{\mathbf{x}\in\mathbf{S}^{d_1-k+1}}\left\|\mathbf{V}_2^{\mathrm{T}}\mathbf{V}_1\mathbf{x}\right\|^2\\ &= \max_{\mathbf{Q}\in\mathbb{R}^{d_1\times(d_1-k+1)}}\min_{\substack{\|\mathbf{y}\|=1\\ \mathbf{y}\in\mathcal{C}(\mathbf{V}_1\mathbf{Q})}}\left\|\mathbf{P}_{\mathcal{Y}_2^\perp}(\mathbf{y})\right\|^2\\ &\ge \min_{\substack{\|\mathbf{y}\|=1\\ \mathbf{y}\in\mathcal{C}(\mathbf{V}_1\mathbf{Q}_l)}}\left\|\mathbf{P}_{\mathcal{Y}_2^\perp}(\mathbf{y})\right\|^2\\ &= \min_{\substack{\|\mathbf{y}\|=1\\ \mathbf{y}\in\mathcal{Y}_{1,k:d_1}}}\left\|\mathbf{P}_{\mathcal{Y}_2^\perp}(\mathbf{y})\right\|^2,\end{aligned}$$

where $\mathbf{Q}_l=[\mathbf{V}_1^{\mathrm{T}}\mathbf{a}_{1,k},\cdots,\mathbf{V}_1^{\mathrm{T}}\mathbf{a}_{1,d_1}]$, $\mathcal{Y}_{1,k:d_1}:=\operatorname{span}\{\mathbf{a}_{1,k},\cdots,\mathbf{a}_{1,d_1}\}$.

To complete the proof, it suffices to prove the following Lemma.

Lemma 1. (The Bound of Maximum and Minimum) Suppose $\mathcal{X}_1$, $\mathcal{X}_2\subset\mathbb{R}^N$ are two subspaces with dimension $d_1\le d_2$, respectively. Principal angles between them are denoted by $\theta_1,\cdots,\theta_{d_1}$. After Gaussian random projection: $\mathcal{X}_l\xrightarrow{\boldsymbol{\Phi}}\mathcal{Y}_l, l=1,2$, which is defined in Notation 1 in [1], we can get $\mathcal{Y}_1,\mathcal{Y}_2\subset\mathbb{R}^n$, and principal angles between them are denoted by $\psi_1,\cdots,\psi_{d_1}$. Denote $\mathbf{V}_2$ as the orthonormal basis of $\mathcal{Y}_2$, $\overline{\sin^2\psi_k}$ is defined in (1). Then there exist constants c_1,c_2 depending only on ε, such that for any $n\ge c_1d_2$, each of following inequalities

$$\max_{\substack{\|\mathbf{y}\|=1\\ \mathbf{y}\in\mathcal{Y}_{1,1:k}}}\left\|\mathbf{P}_{\mathcal{Y}_2^\perp}(\mathbf{y})\right\|^2 \le \overline{\sin^2\psi_k}+\varepsilon\sin^2\theta_k, \tag{9}$$

$$\min_{\substack{\|\mathbf{y}\|=1\\ \mathbf{y}\in\mathcal{Y}_{1,k:d_1}}}\left\|\mathbf{P}_{\mathcal{Y}_2^\perp}(\mathbf{y})\right\|^2 \ge \overline{\sin^2\psi_k}-\varepsilon\sin^2\theta_k. \tag{10}$$

holds with probability at least

$$1-\frac{1}{2}\mathrm{e}^{-c_2n}. \tag{11}$$

Proof. The detailed proof of the bound in (9) is postponed to the next section. The proof of the bound in (10) is similar to that of (9) and is sketched in section 4. □

3 Proof of (9) in Lemma 1

3.1 Problem Transformation

Now we will first express $\mathbf{P}_{\mathcal{Y}_2^\perp}(\mathbf{y})$ by a basis $\{\mathbf{a}_{1,i}^\perp\}_{i=1}^k$ obtained from $\{\mathbf{a}_{1,i}\}_{i=1}^k$ with coefficient vector $\mathbf{b}$. Then for the convenience of estimating, the restriction of $\|\mathbf{y}\|=1$ is relaxed to a restriction on $\|\mathbf{b}\|$.

Recalling the definition of $\bar{\mathbf{A}}_{1,1:k}$ and $\bar{\mathbf{a}}_{1,k}$ in section 1, and noticing that $\bar{\mathbf{A}}_{1,1:k}$ is a basis of $\mathcal{Y}_{1,1:k}$, $\mathbf{y}$ can be spanned by the columns of $\bar{\mathbf{A}}_{1,1:k}$ as

$$\mathbf{y} = \sum_{i=1}^{k} b_i \bar{\mathbf{a}}_{1,i} = \bar{\mathbf{A}}_{1,1:k}\mathbf{b}, \tag{12}$$

where $\mathbf{b} = (b_i)$ denotes the coefficient vector. Define ϕ_i as the principal angel between 1-dimensional subspace $\mathcal{C}(\bar{\mathbf{a}}_{1,i})$ and $\mathcal{Y}_2$, then $\bar{\mathbf{a}}_{1,i}$ can be decomposed as the projections onto $\mathcal{Y}_2$ and its orthogonal complement space $\mathcal{Y}_2^{\perp}$, that is

$$\bar{\mathbf{a}}_{1,i} = \cos\phi_i\ \bar{\mathbf{a}}_{1,i}^{\|} + \sin\phi_i\ \bar{\mathbf{a}}_{1,i}^{\perp}, \tag{13}$$

where $\bar{\mathbf{a}}_{1,i}^{\|}$ is the unit vector along the projection of $\bar{\mathbf{a}}_{1,i}$ onto $\mathcal{Y}_2$, $\bar{\mathbf{a}}_{1,i}^{\perp}$ is, as defined in section 1, the unit vector along the projection of $\bar{\mathbf{a}}_{1,i}$ onto $\mathcal{Y}_2^{\perp}$. Plugging (13) into (12) and making comparison with (4), we can get

$$\mathbf{P}_{\mathcal{Y}_2^{\perp}}(\mathbf{y}) = \sum_{i=1}^{k} b_i \sin\phi_i \bar{\mathbf{a}}_{1,i}^{\perp}. \tag{14}$$

Now we have got the expression of $\mathbf{P}_{\mathcal{Y}_2^{\perp}}(\mathbf{y})$ under the basis $\{\bar{\mathbf{a}}_{1,i}^{\perp}\}$. We will next loose the condition of $\|\mathbf{y}\| = 1$.

According to (12), the condition $\|\mathbf{y}\| = 1$ in maximum problem in the LHS of (9) can be loosen to the condition on the norm of $\mathbf{b}$.

$$1 = \|\mathbf{y}\|^2 = \left\|\bar{\mathbf{A}}_{1,1:k}\mathbf{b}\right\|^2 \geq \sigma_k^2\left(\bar{\mathbf{A}}_{1,1:k}\right)\|\mathbf{b}\|^2,$$

that is

$$\|\mathbf{b}\|^2 \leq \frac{1}{\sigma_k^2\left(\bar{\mathbf{A}}_{1,1:k}\right)} =: b_u. \tag{15}$$

Considering that the singular values of normalized standard Gaussian random matrix are all very close to 1 (one can refer to the section 3 in [2] for more details), this inequality is not very loose.

Inserting (14) and (15) into the LHS of (9), we can get

$$\begin{aligned}
\max_{\substack{\|\mathbf{y}\|=1 \\ \mathbf{y}\in\mathcal{Y}_{1,1:k}}} \left\|\mathbf{P}_{\mathcal{Y}_2^{\perp}}(\mathbf{y})\right\|^2 &= \max_{\substack{\|\mathbf{y}\|=1 \\ \mathbf{y}\in\mathcal{Y}_{1,1:k}}} \left\|\sum_{i=1}^{k} b_i \sin\phi_i \bar{\mathbf{a}}_{1,i}^{\perp}\right\|^2 \\
&\leq \max_{\|\mathbf{b}\|^2 \leq b_u} \left\|\sum_{i=1}^{k} b_i \sin\phi_i \bar{\mathbf{a}}_{1,i}^{\perp}\right\|^2 \\
&= \max_{\|\mathbf{b}\|^2 = b_u} \left\|\sum_{i=1}^{k} b_i \sin\phi_i \bar{\mathbf{a}}_{1,i}^{\perp}\right\|^2.
\end{aligned} \tag{16}$$

To establish (9), it suffices to prove that under the condition in Lemma 1, with probability in (11), we have

$$\max_{\|\mathbf{b}\|^2=b_u} \left\| \sum_{i=1}^{k} b_i \sin\phi_i \bar{\mathbf{a}}_{1,i}^{\perp} \right\|^2 \leq \overline{\sin^2\psi}_k + \varepsilon \sin^2\theta_k. \tag{17}$$

3.2 Proof of (17)

Recalling the definition of $\bar{\mathbf{A}}_{1,1:k}^{\perp}$ in section 1, we can get

$$\begin{aligned}
\max_{\|\mathbf{b}\|^2=b_u} \left\| \sum_{i=1}^{k} b_i \sin\phi_i \bar{\mathbf{a}}_{1,i}^{\perp} \right\|^2 &\leq \left(\max_{1\leq i\leq k} \sin^2\phi_i \right) \max_{\|\mathbf{b}\|^2=b_u} \left\| \sum_{i=1}^{k} b_i \bar{\mathbf{a}}_{1,i}^{\perp} \right\|^2 \\
&= \left(\max_{1\leq i\leq k} \sin^2\phi_i \right) \max_{\|\mathbf{b}\|^2=b_u} \left\| \bar{\mathbf{A}}_{1,1:k}^{\perp} \mathbf{b} \right\|^2 \\
&\leq \left(\max_{1\leq i\leq k} \sin^2\phi_i \right) \sigma_1^2 \left(\bar{\mathbf{A}}_{1,1:k}^{\perp} \right) b_u \\
&= \left(\max_{1\leq i\leq k} \sin^2\phi_i \right) \frac{\sigma_1^2 \left(\bar{\mathbf{A}}_{1,1:k}^{\perp} \right)}{\sigma_k^2 \left(\bar{\mathbf{A}}_{1,1:k} \right)} \\
&= \max_{1\leq i\leq k} \sin^2\phi_i + \left(\max_{1\leq i\leq k} \sin^2\phi_i \right) \left(\frac{\sigma_1^2 \left(\bar{\mathbf{A}}_{1,1:k}^{\perp} \right)}{\sigma_k^2 \left(\bar{\mathbf{A}}_{1,1:k} \right)} - 1 \right),
\end{aligned} \tag{18}$$

where the forth line uses the definition of b_u in (15). We then use the following two Lemmas, that is Lemma 2 and Lemma 4, to establish (17).

Lemma 2. Using the same notations in Theorem 1 and section 1, denote ϕ_i as the principal angle between 1-dimensional subspace $\mathcal{C}(\mathbf{a}_{1,i})$ and $\mathcal{Y}_2$, then there exist positive constants c_1,c_2 depending only on ε, such that for any $n > c_1 d_2$, each of the following inequalities

$$\max_{1\leq i\leq k} \sin^2\phi_i \leq \overline{\sin^2\psi}_k + \varepsilon \sin^2\theta_k = \left(1 - \frac{d_2}{n} + \varepsilon\right) \sin^2\theta_k, \tag{19}$$

$$\min_{k\leq i\leq d_1} \sin^2\phi_i \geq \overline{\sin^2\psi}_k - \varepsilon \sin^2\theta_k = \left(1 - \frac{d_2}{n} - \varepsilon\right) \sin^2\theta_k. \tag{20}$$

holds with probability at least $1 - \mathrm{e}^{-c_2 n}$.

Proof. We will use Lemma 3 in [1], of which the proof is included in [2], to complete the proof. For completeness, this lemma is written as below.

Lemma 3. Using the same notations in Lemma 2, there exist positive constants c_1,c_2 depending only on ε, such that for any $n > c_1 d_2$, for each $1 \leq k \leq d_1$,

$$1 - \overline{\sin^2\psi}_k - \varepsilon \sin^2\theta_k \leq \cos^2\phi_k \leq 1 - \overline{\sin^2\psi}_k + \varepsilon \sin^2\theta_k \tag{21}$$

holds with probability at least $1 - \mathrm{e}^{-c_2 n}$.

It is easy to derive from (21) that for each $1 \le k \le d_1$,

$$\overline{\sin^2 \psi}_k - \varepsilon \sin^2 \theta_k \le \sin^2 \phi_k \le \overline{\sin^2 \psi}_k + \varepsilon \sin^2 \theta_k. \tag{22}$$

According to the definition of $\overline{\sin^2 \psi}_k$ in (1), $\overline{\sin^2 \psi}_k \pm \varepsilon \sin^2 \theta_k = (1 - d_2/n \pm \varepsilon) \sin^2 \theta_k$ increases monotonically with $\sin \theta_k \in [0, 1]$. Considering that $0 \le \sin \theta_1 \le \cdots \le \sin \theta_{d_1} \le 1$, two inequalities in Lemma 2 can be derived from two inequalities in (22), respectively, with probability at least $1 - d_1 \mathrm{e}^{-c_2 n}$. Redefine $c_1 := \max\{c_1, 2/c_2\}$, $c_2 := c_2 - 1/c_1$, then when $n \ge c_1 d_2$, the probability is at least $1 - \mathrm{e}^{-c_2 n}$. According to the definition of $\overline{\sin^2 \psi}_k$ in (1), we can get equations in (19) and (20). □

Inequality (19) contributes to the proof of (9). It indicates that the first item $\max_{1 \le i \le k} \sin^2 \phi_i$ in the RHS of (18) is close to $\overline{\sin^2 \psi}_k$.

Lemma 4. Using the same notations in Theorem 1 and section 1, there exist positive constants c_1,c_2 depending only on ε, such that for any $n > c_1 d_2$, each of the following inequalities

$$\frac{\sigma_1^2 \left(\bar{\mathbf{A}}_{1,1:k}^{\perp} \right)}{\sigma_k^2 \left(\bar{\mathbf{A}}_{1,1:k} \right)} - 1 \le \varepsilon, \tag{23}$$

$$1 - \frac{\sigma_{d_1-k+1}^2 \left(\bar{\mathbf{A}}_{1,k:d_1}^{\perp} \right)}{\sigma_1^2 \left(\bar{\mathbf{A}}_{1,k:d_1} \right)} \le \varepsilon \tag{24}$$

holds with probability at least $1 - \mathrm{e}^{-c_2 n}$.

Proof. Based on the definition of standard Gaussian random matrix in Definition 1, we complete the proof through the following two Lemmas, of which the proofs are postponed to section 5 and section 6, respectively.

Lemma 5. Matrix $\bar{\mathbf{A}}_{1,1:k}^{\perp}$ has the same singular values as some normalized standard Gaussian random matrix in $\mathbb{R}^{(n-d_2)\times k}$. Matrix $\bar{\mathbf{A}}_{1,k:d_1}^{\perp}$has the same singular values as some normalized standard Gaussian random matrix in $\mathbb{R}^{(n-d_2)\times(d_1-k+1)}$.

Lemma 6. Assume that $\bar{\mathbf{A}} \in \mathbb{R}^{n\times k} \bar{\mathbf{B}} \in \mathbb{R}^{(n-d_2)\times k}$ are both normalized standard Gaussian random matrix, then there exist positive constants c_1,c_2 depending only on ε, such that for any $n > c_1 k$, each of the following inequalities

$$\frac{\sigma_1^2(\bar{\mathbf{B}})}{\sigma_k^2(\bar{\mathbf{A}})} - 1 \le \varepsilon, \tag{25}$$

$$1 - \frac{\sigma_k^2(\bar{\mathbf{B}})}{\sigma_1^2(\bar{\mathbf{A}})} \le \varepsilon \tag{26}$$

holds with probability at least $1 - \mathrm{e}^{-c_2 n}$.

According to Lemma 6, noticing the conclusion in Lemma 5 and the fact that $d_2 \ge d_1 \ge \max\{k, d_1 - k + 1\}$, we can complete the proof. □

Note that inequality (23) has appeared in [1] and (24) is newly devised.

Inequality (25) contributes to the proof of (23), it indicates that the second item in the RHS of (18) is a small quantity.

Plugging (19) and (23) into (18), there exist positive constants $c_{1,1}$,$c_{2,1}$ depending only on ε_1, $c_{1,2}$,$c_{2,2}$ depending only on ε_2, such that for any $n > \max\{c_{1,1}, c_{1,2}\}d_2$,

$$\begin{aligned}\max_{\|\mathbf{b}\|^2=b_u} \left\|\sum_{i=1}^{k} b_i \sin\phi_i \bar{\mathbf{a}}_{1,i}^{\perp}\right\|^2 &\leq \overline{\sin^2\psi}_k + \varepsilon_1 \sin^2\theta_k + \varepsilon_2 \left(1 - \frac{d_2}{n} + \varepsilon_1\right) \sin^2\theta_k \\ &\leq \overline{\sin^2\psi}_k + (\varepsilon_1 + \varepsilon_2 + \varepsilon_1\varepsilon_2) \sin^2\theta_k. \end{aligned} \tag{27}$$

holds with probability at least $1 - \mathrm{e}^{-c_{2,1}n} - \mathrm{e}^{-c_{2,2}n}$. Let $\varepsilon_1 := \varepsilon_2 := \varepsilon/3 < 1$, take $c_1 := \max\{c_{1,1}, c_{1,2}, \ln 5/\min\{c_{2,1}, c_{2,2}\}\}$, $c_2 := \min\{c_{2,1}, c_{2,2}\} - \ln 4/c_1$. Noticing that the value of c_1 indicates that $c_2 > 0$, we can close the proof.

4 Proof of (10) in Lemma 1

For $\mathbf{y} \in \mathcal{Y}_{1,k:d_1}$, following the same approach we used in section 3.1, we could derive step by step the counterparts of (14), (15) and (16), respectively, as

$$\mathbf{P}_{\mathcal{Y}_2^{\perp}}(\mathbf{y}) = \sum_{i=k}^{d_1} b_i \sin\phi_i \bar{\mathbf{a}}_{1,i}^{\perp},$$

$$\|\mathbf{b}\|^2 \geq \frac{1}{\sigma_1^2\left(\bar{\mathbf{A}}_{1,k:d_1}\right)} =: b_l,$$

and

$$\min_{\substack{\|\mathbf{y}\|=1 \\ \mathbf{y}\in\mathcal{Y}_{1,k:d_1}}} \left\|\mathbf{P}_{\mathcal{Y}_2^{\perp}}(\mathbf{y})\right\|^2 = \min_{\|\mathbf{b}\|^2=b_l} \left\|\sum_{i=k}^{d_1} b_i \sin\phi_i \bar{\mathbf{a}}_{1,i}^{\perp}\right\|^2,$$

where $\bar{\mathbf{A}}_{1,k:d_1}$ is defined in section 1. Combing the above two equalities and one inequality together and following the approach when deriving (18), we could get

$$\begin{aligned}\min_{\substack{\|\mathbf{y}\|=1 \\ \mathbf{y}\in\mathcal{Y}_{1,k:d_1}}} \left\|\mathbf{P}_{\mathcal{Y}_2^{\perp}}(\mathbf{y})\right\|^2 &\geq \left(\min_{k\leq i\leq d_1} \sin^2\phi_i\right) \frac{\sigma_{d_1-k+1}^2\left(\bar{\mathbf{A}}_{1,k:d_1}^{\perp}\right)}{\sigma_1^2\left(\bar{\mathbf{A}}_{1,k:d_1}\right)} \\ &\geq \min_{k\leq i\leq d_1} \sin^2\phi_i - \left(\min_{k\leq i\leq d_1} \sin^2\phi_i\right) \left(1 - \frac{\sigma_{d_1-k+1}^2\left(\bar{\mathbf{A}}_{1,k:d_1}^{\perp}\right)}{\sigma_1^2\left(\bar{\mathbf{A}}_{1,k:d_1}\right)}\right),\end{aligned}$$

where $\bar{\mathbf{A}}_{1,k:d_1}^{\perp}$ is defined in section 1. Using (20) in Lemma 2 and (24) in Lemma 4, we can get

$$\min_{\substack{\|\mathbf{y}\|=1 \\ \mathbf{y}\in\mathcal{Y}_{1,k:d_1}}} \left\|\mathbf{P}_{\mathcal{Y}_2^{\perp}}(\mathbf{y})\right\|^2 \geq \overline{\sin^2\psi}_k - (\varepsilon_1 + \varepsilon_2 - \varepsilon_1\varepsilon_2) \sin^2\theta_k.$$

Letting $\varepsilon_1 := \varepsilon_2 := \varepsilon/2$ and taking $c_1 := \max\{c_{1,1}, c_{1,2}, \ln 5/\min\{c_{2,1}, c_{2,2}\}\}$, $c_2 := \min\{c_{2,1}, c_{2,2}\} - \ln 4/c_1$, we can close the proof.

5 Proof of Lemma 5

We will first state the independence between column-spanned subspaces of random matrices, and then prove Lemma 5 through two Lemmas.

Definition 2. (Independence of Subspaces) Two subspaces are independent, if and only if they are spanned by the columns of two independent random matrices, respectively, where two random matrices are independent if and only if any two columns of them are independent.

For subspaces spanned by Gaussian random matrices, the independence of them can be verified through the following Lemma.

Lemma 7. [2] Assume $\mathbf{U}$ and $\mathbf{V}$ are two matrices satisfying $\mathbf{U}^{\mathrm{T}}\mathbf{V} = \mathbf{0}$, and $\mathbf{\Phi}$ is a standard Gaussian random matrix. Then column spaces $\mathcal{C}(\mathbf{\Phi U})$ and $\mathcal{C}(\mathbf{\Phi V})$ are independent.

Furthermore, we have the following lemma about Gaussian random projection.

Lemma 8. [2] Let $\mathbf{H}_1 \in \mathbb{R}^{n\times k_1}$ and $\mathbf{H}_2 \in \mathbb{R}^{n\times k_2}$, $k_1 \le k_2$ be two standard Gaussian random matrices. We denote $\mathbf{Q}_2$ as an orthonormal basis of the column space $\mathcal{C}(\mathbf{H}_2)$. The projection of $\mathbf{H}_1 := [\mathbf{h}_{1,1}, \cdots, \mathbf{h}_{1,k_1}]$ onto $\mathcal{C}(\mathbf{H}_2)$ is denoted by $\mathbf{B}_1 := [\mathbf{b}_{1,1}, \cdots, \mathbf{b}_{1,k_1}]$, i.e., $\mathbf{b}_{1,i}$ is the projection of $\mathbf{h}_{1,i}$. If $\mathbf{H}_1$ and $\mathbf{H}_2$ are independent, we have $\mathbf{B}_1 = \mathbf{Q}_2\mathbf{\Omega}$, where $\mathbf{\Omega} \in \mathbb{R}^{k_2\times k_1}$ is a standard Gaussian random matrix.

Project $\mathbf{u}_{1,k}$, $k = 1, \cdots, d_1$ onto $\mathcal{Y}_2$ and its orthonormal complement space $\mathcal{Y}_2^\perp$ to get $\mathbf{u}_{2,k}$ and $\mathbf{u}_{0,k}$, respectively, as follows

$$\mathbf{u}_{1,k} = \cos\theta_k\ \mathbf{u}_{2,k} + \sin\theta_k\ \mathbf{u}_{0,k},$$

where $\mathbf{u}_{0,k}$ is orthogonal to $\mathcal{Y}_2$ and its l_2 norm equals 1. After random projection, denote $\mathbf{a}_{0,k} := \mathbf{\Phi}\mathbf{u}_{0,k}$, then

$$\mathbf{a}_{1,k} = \cos\theta_k\ \mathbf{a}_{2,k} + \sin\theta_k\ \mathbf{a}_{0,k}. \tag{28}$$

Project both sides of (28) onto $\mathcal{Y}_2^\perp$. Considering that $\mathbf{a}_{2,k}$ lies in $\mathcal{Y}_2$, the projection of $\mathbf{a}_{1,k}$ is along the projection of $\mathbf{a}_{0,k}$. Denote the normalized projection of $\mathbf{a}_{0,k}$ onto $\mathcal{Y}_2^\perp$ as $\bar{\mathbf{a}}_{0,k}^\perp$. Recalling the definition of $\bar{\mathbf{a}}_{1,k}^\perp$, we have

$$\bar{\mathbf{a}}_{1,k}^\perp = \bar{\mathbf{a}}_{0,k}^\perp.$$

Denote

$$\mathbf{U}_{0,1:k} = [\mathbf{u}_{0,1}, \cdots, \mathbf{u}_{0,k}], \quad \mathbf{A}_{0,1:k}^\perp := \left[\mathbf{a}_{0,1}^\perp, \cdots, \mathbf{a}_{0,k}^\perp\right], \quad \bar{\mathbf{A}}_{0,1:k}^\perp := \left[\bar{\mathbf{a}}_{0,1}^\perp, \cdots, \bar{\mathbf{a}}_{0,k}^\perp\right],$$

then

$$\bar{\mathbf{A}}_{1,1:k}^\perp = \bar{\mathbf{A}}_{0,1:k}^\perp. \tag{29}$$

So it suffices to prove that $\bar{\mathbf{A}}_{0,1:k}^{\perp}$ has the same singular values as some normalized standard Gaussian random matrix in $\mathbb{R}^{(n-d_2)\times k}$. Because $\mathbf{u}_{0,k}$ is orthogonal to $\mathcal{Y}_2$, we have $\mathbf{U}_{0,1:k}^{\mathrm{T}}\mathbf{U}_2 = \mathbf{0}$. According to Lemma 7, the column space of $\mathbf{A}_{0,1:k} := \mathbf{\Phi}\mathbf{U}_{0,1:k}$ is independent with the column space of $\mathbf{A}_2 = \mathbf{\Phi}\mathbf{U}_2$, that is $\mathcal{Y}_2$, and thus independent with its orthogonal complement space $\mathcal{Y}_2^{\perp}$. Denote $\mathbf{V}_2^{\perp}$ as an arbitrary orthonormal basis matrix for $\mathcal{Y}_2^{\perp}$, then according to Lemma 8, the projection of $\mathbf{A}_{0,1:k}$ onto this subspace can be written as $\mathbf{A}_{0,1:k}^{\perp} := \mathbf{V}_2^{\perp}\mathbf{\Omega}$, where $\mathbf{\Omega} \in \mathbb{R}^{(n-d_2)\times k}$ is a standard Gaussian random matrix. Since that $\mathbf{V}_2^{\perp}$ is a column-orthonormal matrix, $\mathbf{A}_{0,1:k}^{\perp}$ has the same singular values as $\mathbf{\Omega}$. $\bar{\mathbf{A}}_{0,1:k}^{\perp}$ is the column-normalized result of $\mathbf{A}_{0,1:k}^{\perp}$, so it has the same singular values as some normalized standard Gaussian random matrix in $\mathbb{R}^{(n-d_2)\times k}$. Now we have proven the first part of Lemma 5. For the second part, in a similar way, we can also have

$$\bar{\mathbf{A}}_{1,k:d_1}^{\perp} = \bar{\mathbf{A}}_{0,k:d_1}^{\perp}.$$

Thus, we complete the proof.

6 Proof of Lemma 6

In this section, we will first introduce some useful Lemmas and then give the detailed proof of (25). (26) can be reached in a similar way.

6.1 Useful Lemmas

The first Lemma is about the singular value of normalized standard Gaussian random matrix.

Lemma 9. [2] Assume that $\bar{\mathbf{A}} \in \mathbb{R}^{n\times k}$ is a normalized standard Gaussian random matrix, then there exist positive constants $c_{1,1}$, $c_{2,1}$, $c_{1,2}$, $c_{2,2}$ depending only on ε, such that

$$\mathbb{P}\left(\sigma_1^2(\bar{\mathbf{A}}) < 1+\varepsilon\right) \geq 1 - \mathrm{e}^{-c_{2,1}n}, \quad \forall n \geq c_{1,1}k, \tag{30}$$

$$\mathbb{P}\left(\sigma_k^2(\bar{\mathbf{A}}) > 1-\varepsilon\right) \geq 1 - \mathrm{e}^{-c_{2,2}n}, \quad \forall n \geq c_{1,2}k. \tag{31}$$

The next Lemma is to verify that the function satisfying certain condition can be written as a single exponential function.

Lemma 10. [2] Given

$$f(\varepsilon, n, \tau) = \frac{1}{K}\sum_{k=1}^{K} a_k(\varepsilon, n)\mathrm{e}^{-g_k(\varepsilon,n,\tau)},$$

if for all k, it holds that

$$h_k(\varepsilon) := \lim_{\tau\to 0}\lim_{n\to\infty}\frac{g_k(\varepsilon,n,\tau)}{n} > 0, \tag{32}$$

$$b_k(\varepsilon) := \lim_{n\to\infty}\frac{\ln a_k(\varepsilon,n)}{n} < h_k(\varepsilon), \tag{33}$$

then there exist positive constants n_0, τ_0, c_2 depending only on ε, such that for any $n > n_0$, $\tau < \tau_0$, it satisfies that $f(\varepsilon, n, \tau) < \mathrm{e}^{-c_2 n}$.

6.2 Proof of (25)

According to Lemma 9, there exist positive constants $c_{1,1}$, $c_{2,1}$, $c_{1,2}$, $c_{2,2}$ depending only on ε_1, such that for any $n > \max\{c_{1,1}, c_{1,2}\}k$,

$$\frac{\sigma_1^2(\bar{\mathbf{B}})}{\sigma_k^2(\bar{\mathbf{A}})} - 1 \le \frac{1+\varepsilon_1}{1-\varepsilon_1} - 1 = \frac{2\varepsilon_1}{1-\varepsilon_1} \le 4\varepsilon_1 \tag{34}$$

holds with probability at least $1 - \mathrm{e}^{-c_{2,1}(n-d_2)} - \mathrm{e}^{-c_{2,2}n}$, where the last inequality holds for $\varepsilon_1 \le 1/2$. Let $\varepsilon := 4\varepsilon_1$, $c_1 := \max\{c_{1,1}, c_{1,2}\}$. Then for any $n > c_1 k$, (25) holds with probability at least $1 - \mathrm{e}^{-c_{2,1}(n-d_2)} - \mathrm{e}^{-c_{2,2}n}$. To establish (25) in Lemma 6, it suffices to prove that there exists positive constant c_2, such that $\mathrm{e}^{-c_{2,1}(n-d_2)} + \mathrm{e}^{-c_{2,2}n} \le \mathrm{e}^{-c_2 n}$. According to Lemma 10, let $\tau := d_2/n$, $a_1 := 2$, $g_1(\varepsilon, n, \tau) := c_{2,1}(n - d_2)$, then

$$\begin{aligned} h_1(\varepsilon) &:= \lim_{\tau\to 0}\lim_{n\to\infty} \frac{g_1(\varepsilon,n,\tau)}{n} = \lim_{\tau\to 0} c_{2,1}(1-\tau) = c_{2,1} > 0, \\ b_1(\varepsilon) &:= \lim_{n\to\infty} \frac{\ln a_1}{n} = 0 < h_1(\varepsilon) \end{aligned}$$

satisfies condition (32) and (33), respectively. In a similar way, the item $\mathrm{e}^{-c_{2,2}n}$ also satisfies the above two conditions. So there exist constants n_0, τ_0, c_2 such that for any $n > n_0$, $d_2/n < \tau_0$, that is $n > d_2/\tau_0$, we have $\mathrm{e}^{-c_{2,1}(n-d_2)} + \mathrm{e}^{-c_{2,2}n} \le \mathrm{e}^{-c_2 n}$. Redefine $c_1 := \max\{c_{1,1}, c_{1,2}, n_0, 1/\tau_0\}$, we can close the proof.

6.3 Proof of (26)

According to Lemma 9, there exist positive constants $c_{1,1}$, $c_{1,2}$, $c_{2,1}$, $c_{2,2}$ depending only on ε_2, such that for any $n > \max\{c_{1,1}, c_{1,2}\}k$,

$$1 - \frac{\sigma_k^2(\bar{\mathbf{B}})}{\sigma_1^2(\bar{\mathbf{A}})} \le 1 - \frac{1-\varepsilon_2}{1+\varepsilon_2} \le 2\varepsilon_2$$

holds with probability at least $1 - \mathrm{e}^{-c_{2,1}(n-d_2)} - \mathrm{e}^{-c_{2,2}n}$. Let $\varepsilon_2 := \varepsilon/2$, and then applying Lemma 10 we can complete the proof of (26).

References

[1] Y. Jiao, X. Shen, G. Li, and Y. Gu, "Subspace principal angle preserving property of gaussian random projection," submitted to *IEEE Data Science Workshop*, 2018.

[2] G. Li, Q. Liu, and Y. Gu, "Rigorous restricted isometry property of low-dimensional subspaces," *arXiv:1801.10058*, 2018.