0.1. Vector Hoeffding inequality

Theorem 1. Let $I\!\!E X_i = 0$, $\|X_i\| < c_i/2$, then for independent random variables

$$I\!\!P\left(\left\|\sum_{i=1}^n X_i\right\| > t\right) \le \exp\left(-\frac{(t-v)^2}{2v^2}\right), \quad t \ge v,$$

$$v^2 = \frac{1}{4}\sum_{i=1}^n c_i^2.$$

$$I\!\!P\left(\left\|\sum_{i=1}^n X_i\right\| > v(1+\sqrt{2x})\right) \le e^{-x}.$$

Proof. Bounded difference inequality state with $|Z(\ldots, X_i, \ldots) - Z(\ldots, X_i', \ldots)| \le c_i$ that

$$I\!\!P(Z - I\!\!E Z > t) \le \exp\left(-\frac{2t^2}{\sum_i c_i^2}\right).$$

$$I\!\!P\left(\left\|\sum_{i=1}^n X_i\right\| - I\!\!E\left\|\sum_{i=1}^n X_i\right\| > t - I\!\!E\left\|\sum_{i=1}^n X_i\right\|\right) \le \exp\left(-\frac{\left(t - I\!\!E\left\|\sum_{i=1}^n X_i\right\|\right)^2}{2v^2}\right),$$

$$I\!\!E\left\|\sum_{i=1}^n X_i\right\| \le \sqrt{I\!\!E\left\|\sum_{i=1}^n X_i\right\|^2} \le \sqrt{\sum_{i=1}^n I\!\!E\|X_i\|^2} \le v.$$

□

0.2. Matrix Bernstein inequality

Theorem 2 (Master bound)**.** Assume that $\boldsymbol{S}_1, \ldots, \boldsymbol{S}_n$ are independent Hermitian matrices of the same size and $\boldsymbol{Z} = \sum_{i=1}^n \boldsymbol{S}_i$. Then

$$I\!\!E\lambda_{\max}(\boldsymbol{Z}) \le \inf_{\theta>0} \frac{1}{\theta} \log \operatorname{tr} \exp\left(\sum_{i=1}^n \log I\!\!E \mathrm{e}^{\theta \boldsymbol{S}_i}\right), \tag{1}$$

$$I\!\!P\{\lambda_{\max}(\boldsymbol{Z}) \ge z\} \le \inf_{\theta>0} \mathrm{e}^{-\theta z} \operatorname{tr} \exp\left(\sum_{i=1}^n \log I\!\!E \mathrm{e}^{\theta \boldsymbol{S}_i}\right). \tag{2}$$

Proof. By the Markov inequality

$$I\!\!P\{\lambda_{\max}(\boldsymbol{Z}) \ge z\} \le \inf_{\theta} \mathrm{e}^{-\theta z} I\!\!E \exp(\theta \lambda_{\max}(\boldsymbol{Z})).$$

Recall the spectral mapping theorem: for any function $f\colon I\!\!R \to I\!\!R$ and Hermitian matrix A eigenvalues of $f(A)$ are equal to eigenvalues of A. Thus

$$\exp(\theta\lambda_{\max}(\boldsymbol{Z})) = \exp(\lambda_{\max}(\theta\boldsymbol{Z})) = \lambda_{\max}\big(\exp(\theta\boldsymbol{Z})\big) \le \operatorname{tr} \mathrm{e}^{\theta\boldsymbol{Z}}.$$

Therefore,

$$I\!\!P\{\lambda_{\max}(\boldsymbol{Z}) \ge z\} \le \inf_{\theta} \mathrm{e}^{-\theta z} I\!\!E \operatorname{tr} \exp(\theta\boldsymbol{Z}), \tag{3}$$

and (**??**) follows.

To prove (**??**) fix θ. Using the spectral mapping theorem one can get that

$$I\!\!E\lambda_{\max}(\boldsymbol{Z}) = \frac{1}{\theta} I\!\!E\lambda_{\max}(\theta\boldsymbol{Z}) = \frac{1}{\theta}\log I\!\!E\exp\big(\lambda_{\max}(\theta\boldsymbol{Z})\big) = \frac{1}{\theta}\log I\!\!E\lambda_{\max}\big(\exp(\theta\boldsymbol{Z})\big).$$

Thus we get

$$\mathbb{E}\lambda_{\max}(\boldsymbol{Z}) \le \frac{1}{\theta} \log \operatorname{tr} \mathbb{E} \exp(\theta \boldsymbol{Z}). \tag{4}$$

The final step in proving the master inequalities is to bound from above $\mathbb{E} \operatorname{tr} \exp\left(\sum_{i=1}^n \boldsymbol{S}_i\right)$. To do this we use Jensen's inequality for the convex function $\operatorname{tr}\exp(H + \log(X))$ (in matrix X), where H is deterministic Hermitian matrix. For a random Hermitian matrix X one can write

$$\mathbb{E} \operatorname{tr} \exp(H + X) = \mathbb{E} \operatorname{tr} \exp(H + \log \mathrm{e}^X) \le \operatorname{tr} \exp(H + \log \mathbb{E} \mathrm{e}^X). \tag{5}$$

Convexity of function $(\operatorname{tr}\exp(H + \log(X)))$ is followed from

$$\operatorname{tr}\exp(H + \log(X)) = \max_{Y \succ 0} [\operatorname{tr}(YH) - (D(Y;X) - \operatorname{tr} X)],$$

where $D(Y;X)$ is relative entropy

$$D(Y;X) = \phi(X) - [\phi(Y) + \langle \nabla\phi(Y), X - Y \rangle], \quad \phi(X) = \operatorname{tr}(X \log X)$$

due to the partial maximum and $D(Y;X)$ are concave functions.

Denote by $\mathbb{E}_i$ the conditional expectation with respect to random matrix X_i. To bound $\mathbb{E} \operatorname{tr} \exp\left(\sum_{i=1}^n \boldsymbol{S}_i\right)$ we use (**??**) for the sum of independent Hermitian matrices by taking the conditional expectations with respect to i-th matrix:

$$\begin{aligned}
\mathbb{E} \operatorname{tr} \exp\left(\sum_{i=1}^n \boldsymbol{S}_i\right) &= \mathbb{E}\mathbb{E}_n \operatorname{tr} \exp\left(\sum_{i=1}^{n-1} \boldsymbol{S}_i + \boldsymbol{S}_n\right) \\
&\le \mathbb{E} \operatorname{tr} \exp\left(\sum_{i=1}^{n-1} \boldsymbol{S}_i + \log(\mathbb{E}_n \exp(\boldsymbol{S}_n))\right) \\
&\le \operatorname{tr} \exp\left(\sum_{i=1}^{n} \log \mathbb{E} \mathrm{e}^{\theta \boldsymbol{S}_i}\right).
\end{aligned} \tag{6}$$

To complete the prove of the Master's theorem combine (**??**) and (**??**) with (**??**). □

The same result applied to $-\boldsymbol{Z}$ yields the bound for the operator norm $\|\boldsymbol{Z}\|$:

$$\mathbb{P}\{\|\boldsymbol{Z}\|_{\mathrm{op}} \ge z\} \le 2 \inf_{\theta > 0} \mathrm{e}^{-\theta z} \operatorname{tr} \exp\left(\sum_{i=1}^n \log \mathbb{E} \mathrm{e}^{\theta \boldsymbol{S}_i}\right). \tag{7}$$

Theorem 3 (Bernstein inequality for a sum of random Hermitian matrices)**.** Let $\boldsymbol{Z} = \sum_{i=1}^n \boldsymbol{S}_i$, where $\boldsymbol{S}_i$, $i = 1, \ldots, n$ are independent, random, Hermitian matrices of the dimension $d \times d$ and

$$\lambda_{\max}(\boldsymbol{S}_i) \le R.$$

Denote $\mathrm{v}^2 = \mathrm{v}^2(\boldsymbol{Z}) = \|\mathbb{E}(\boldsymbol{Z}^2)\|_{\mathrm{op}}$. Then

$$\mathbb{E}\lambda_{\max}(\boldsymbol{Z}) \le \sqrt{2\mathrm{v}^2 \log(d)} + \frac{1}{3} R \log(d), \tag{8}$$

$$\mathbb{P}\big\{\lambda_{\max}(\boldsymbol{Z}) \ge z\big\} \le d \exp\left(\frac{-z^2/2}{\mathrm{v}^2 + Rz/3}\right). \tag{9}$$

Proof. Note that

$$\mathrm{v}^2 = \left\| \sum_{i=1}^{n} \mathbb{E} \boldsymbol{S}_i^2 \right\|_{\mathrm{op}} .$$

For the sake of simplicity let $\mathrm{v}^2 = 1$. Denote

$$g(\theta) = \frac{\theta^2/2}{1 - R\theta/3}.$$

Apart the Master inequalities, we use the following lemma:

Lemma. Let $\boldsymbol{Z}$ be a random Hermitian matrix $\mathbb{E}\boldsymbol{Z} = 0$, $\lambda_{\max}(\boldsymbol{Z}) \le R$, then for $0 < \theta < 3/R$ the following inequalities hold

$$\mathbb{E}\mathrm{e}^{\theta \boldsymbol{Z}} \le \exp\left(\frac{\theta^2/2}{1 - R\theta/3} \mathbb{E}(\boldsymbol{Z}^2) \right),$$

$$\log \mathbb{E}\mathrm{e}^{\theta \boldsymbol{Z}} \le \frac{\theta^2/2}{1 - R\theta/3} \mathbb{E}(\boldsymbol{Z}^2).$$

Proof. Decompose the exponent in the following way

$$\mathrm{e}^{\theta \boldsymbol{Z}} = I + \theta \boldsymbol{Z} + (\mathrm{e}^{\theta \boldsymbol{Z}} - \theta \boldsymbol{Z} - I) = I + \theta \boldsymbol{Z} + \boldsymbol{Z} \cdot f(\boldsymbol{Z}) \cdot \boldsymbol{Z},$$

where

$$f(x) = \frac{\mathrm{e}^{\theta x} - \theta x - 1}{x^2}, \quad \text{for} \quad x \ne 0, \quad f(0) = \frac{\theta^2}{2}.$$

One can check that the function $f(x)$ is non-decreasing, thus for $x \le R$, one has $f(x) \le f(R)$. By the matrix transfer rule $f(\boldsymbol{Z}) \le f(R) I$ and

$$\mathbb{E}\mathrm{e}^{\theta \boldsymbol{Z}} = I + f(R) \mathbb{E}\boldsymbol{Z}^2.$$

In order to estimate $f(R)$ use $q! \ge 2 \cdot 3^{q-2}$ to get

$$f(R) = \frac{\mathrm{e}^{\theta R} - \theta R - I}{R^2} = \frac{1}{R^2} \sum_{q=2}^{\infty} \frac{(\theta R)^q}{q!} \le \theta^2 \sum_{q=2}^{\infty} \frac{(R\theta)^{q-2}}{3^{q-2}} = \frac{\theta^2/2}{1 - R\theta/3}.$$

To get the result of the Lemma note that $1 + a \le \mathrm{e}^a$. □

To prove (**??**) and (**??**) we apply the Master inequalities and Lemma 0.2.:

$$\begin{aligned}
\mathbb{E}\lambda_{\max}(\boldsymbol{Z}) &\le \inf_{\theta > 0} \frac{1}{\theta} \log \operatorname{tr} \exp\left(\sum_{i=1}^{n} \log \mathbb{E} \exp(\theta \boldsymbol{S}_i) \right) \\
&\le \inf_{0 < \theta < 3/R} \frac{1}{\theta} \log \operatorname{tr} \exp\left(g(\theta) \sum_{i=1}^{n} \mathbb{E} \boldsymbol{S}_i^2 \right) \\
&\le \inf_{0 < \theta < 3/R} \frac{1}{\theta} \log \operatorname{tr} \exp\left(g(\theta) \mathbb{E} \boldsymbol{Z}^2 \right) \\
&\le \inf_{0 < \theta < 3/R} \frac{1}{\theta} \log d \exp\left(g(\theta) \| \mathbb{E} \boldsymbol{Z}^2 \|_{\mathrm{op}} \right) \\
&\le \inf_{0 < \theta < 3/R} \left\{ \frac{\log(d)}{\theta} + \frac{\theta/2}{1 - R\theta/3} \right\}.
\end{aligned}$$

Minimizing the right hand side in θ one can get (**??**).

The second inequality can be obtained in the same manner:

$$\begin{aligned}
I\!P\{\lambda_{\max}(\boldsymbol{Z}) \geq z\} &\leq \inf_{\theta>0} \mathrm{e}^{-\theta z}\operatorname{tr}\exp\left(\sum_{i=1}^{n}\log I\!E\exp(\theta\boldsymbol{S}_i)\right) \\
&\leq \inf_{0<\theta<3/R} \mathrm{e}^{-\theta z}\operatorname{tr}\exp\left(g(\theta)I\!E\boldsymbol{Z}^2\right) \\
&\leq \inf_{0<\theta<3/R} \mathrm{e}^{-\theta z} d\exp\left(g(\theta)\|I\!E\boldsymbol{Z}^2\|_{\mathrm{op}}\right) \\
&\leq \inf_{0<\theta<3/R} \mathrm{e}^{-\theta z} d\exp\left(g(\theta)\right).
\end{aligned}$$

Here instead of minimizing the right hand side in θ we have used $\theta = z/(1+Rz/3)$. □

Theorem 4 (Bernstein inequality for a sum of random Hermitian matrices)**.** Let $\boldsymbol{Z} = \sum_{i=1}^{n}\boldsymbol{S}_i$, where $\boldsymbol{S}_i$, $i=1,\ldots,n$ are independently distributed random matrices of the size $d_1 \times d_2$ and

$$\|\boldsymbol{S}_i\|_{\mathrm{op}} \leq R.$$

Denote $\mathrm{v}^2 = \mathrm{v}^2(\boldsymbol{Z}) = \max\left\{\|I\!E(\boldsymbol{Z}^*\boldsymbol{Z})\|_{\mathrm{op}}, \|I\!E(\boldsymbol{Z}\boldsymbol{Z}^*)\|_{\mathrm{op}}\right\}$. Then

$$I\!E\|\boldsymbol{Z}\|_{\mathrm{op}} \leq \sqrt{2\mathrm{v}^2\log(d_1+d_2)} + \frac{1}{3}R\log(d),$$

$$I\!P\{\|\boldsymbol{Z}\|_{\mathrm{op}} \geq z\} \leq (d_1+d_2)\exp\left(\frac{-z^2/2}{\mathrm{v}^2+Rz/3}\right).$$

Proof. Use the following hint: define the matrix

$$H(\boldsymbol{Z}) = \begin{pmatrix} 0 & \boldsymbol{Z} \\ \boldsymbol{Z}^* & 0 \end{pmatrix}.$$

It can be easily seen that $\mathrm{v}^2 = \|H(\boldsymbol{Z})^2\|_{\mathrm{op}}$, and $\|\boldsymbol{Z}\|_{\mathrm{op}} = \lambda_{\max}\big(H(\boldsymbol{Z})\big)$, thus applying Theorem 3 to $H(\boldsymbol{Z})$ the statements (**??**) and (**??**) are straightforward. □

Theorem 5 (Bernstein inequality for moment restricted matrices)**.** Suppose that

$$\forall i:\; I\!E\psi^2\left(\frac{\|X_i\|}{M}\right) \leq 1,$$

$$R = M\psi^{-1}\left(\frac{2}{\delta}\frac{nM^2}{\mathrm{v}^2}\right), \qquad \delta \in (0, 2/\psi(1)).$$

Then for $zR \leq (e-1)(1+\delta)\mathrm{v}^2$

$$I\!P\{\|\boldsymbol{Z}\|_{\mathrm{op}} \geq z\} \leq 2p\exp\left\{-\frac{z^2}{2(1+\delta)\mathrm{v}^2+2Rz/3}\right\}$$

A standart case for ψ is $\psi(u) = e^{u^\alpha}-1$, which leads to $R = M\log^{1/\alpha}(\frac{2}{\delta}\frac{nM^2}{\mathrm{v}^2}+1)$.

Proof. According to Master bound one have to estimate $I\!Ee^{\theta\boldsymbol{S}}$ for $\boldsymbol{S}$ in $\boldsymbol{S}_1,\ldots,\boldsymbol{S}_n$. Denote a function

$$f(u) = \frac{e^u-1-u}{u^2}$$

Taylor expansion yields

$$I\!Ee^{\theta\boldsymbol{S}} \leq I_p + \theta^2 I\!E\boldsymbol{S}^2 f(\theta\,\|\boldsymbol{S}\|)$$

$$\log \mathbb{E} e^{\theta \boldsymbol{S}} \le \theta^2 \mathbb{E} \boldsymbol{S}^2 f(\theta \|\boldsymbol{S}\|) \le \theta^2 f(\theta\tau) \mathbb{E} \boldsymbol{S}^2 + I_p \theta^2 \mathbb{E} \|\boldsymbol{S}\|^2 f(\theta \|\boldsymbol{S}\|) I(\|\boldsymbol{S}\| \ge \tau)$$

$$\mathbb{E} \|\boldsymbol{S}\|^2 f(\theta \|\boldsymbol{S}\|) I(\|\boldsymbol{S}\| \ge \tau) \le M^2 \mathbb{E} \psi^2 \left(\frac{\|\boldsymbol{S}\|}{M}\right) \left(\psi\left(\frac{\tau}{M}\right)\right)^{-1} \le M^2 \left(\psi\left(\frac{\tau}{M}\right)\right)^{-1}$$

$$M^2 \left(\psi\left(\frac{R}{M}\right)\right)^{-1} = \frac{\delta \mathrm{v}^2}{2n}$$

$$\log \mathbb{E} e^{\theta \boldsymbol{S}} \le \theta^2 f(\theta R) \mathbb{E} \boldsymbol{S}^2 + I_p \theta^2 \frac{\delta \mathrm{v}^2}{2n}$$

$$\operatorname{tr} \exp\left(\sum_{i=1}^n \log \mathbb{E} \exp(\theta \boldsymbol{S}_i)\right) \le \operatorname{tr} \exp\left(\theta^2 f(\theta R) \mathbb{E} \sum_i \boldsymbol{S}_i^2\right) \exp\left(\theta^2 \frac{\delta \mathrm{v}^2}{2}\right)$$

□

0.3. Matrix deviation bounds

The next result provides a deviation bound for a matrix valued quadratic forms.

Theorem 6 (Deviation bound for matrix quadratic forms)**.** Consider a $p \times n$ matrix $\mathcal{U}$ such that

$$\mathcal{U}\mathcal{U}^\top = I_p.$$

Let the columns $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n \in \mathbb{R}^p$ of the matrix $\mathcal{U}$ satisfy

$$\|\boldsymbol{u}_i\| \le \delta \tag{10}$$

for a fixed constant δ. For a random vector $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^\top$ with independent standard Gaussian components, define

$$\boldsymbol{Z} \stackrel{\text{def}}{=} \mathcal{U} \operatorname{diag}\{\boldsymbol{\varepsilon} \cdot \boldsymbol{\varepsilon} - 1\} \mathcal{U}^\top = \sum_{i=1}^n (\varepsilon_i^2 - 1) \boldsymbol{u}_i \boldsymbol{u}_i^\top.$$

Then

$$\mathbb{P}\Bigl(\|\boldsymbol{Z}\|_{\mathrm{op}} \ge 2\delta \sqrt{\mathtt{x} + \log(2p)} + 2\delta^2 (\mathtt{x} + \log(2p))\Bigr) \le \mathrm{e}^{-\mathtt{x}}.$$

Proof. From the Master inequality (**??**)

$$\mathbb{P}\bigl(\|\boldsymbol{Z}\|_{\mathrm{op}} \ge z\bigr) \le 2 \inf_{\theta > 0} \mathrm{e}^{-\theta z} \operatorname{tr} \exp\Bigl\{\sum_{i=1}^n \log \mathbb{E} \exp(\theta(\varepsilon_i^2 - 1) \boldsymbol{u}_i \boldsymbol{u}_i^\top)\Bigr\}. \tag{11}$$

Now we use the following general fact:

Lemma. If ξ is a random variable and $\mathit{\Pi}$ is a projector in $\mathbb{R}^p$, then

$$\log \mathbb{E} \exp(\xi \mathit{\Pi}) = \log\bigl(\mathbb{E} \mathrm{e}^{\xi}\bigr) \mathit{\Pi}. \tag{12}$$

Proof. The result (**??**) can be easily obtained by applying twice the spectral mapping theorem. □

This result yields, in particular, for any unit vector $\boldsymbol{u} \in \mathbb{R}^p$

$$\log \mathbb{E} \exp\bigl(\xi \boldsymbol{u}\boldsymbol{u}^\top\bigr) = \log\bigl(\mathbb{E} \mathrm{e}^{\xi}\bigr) \boldsymbol{u}\boldsymbol{u}^\top.$$

Moreover, for any vector $\boldsymbol{u} \in I\!\!R^p$, the normalized product $\boldsymbol{u}\boldsymbol{u}^\top / \|\boldsymbol{u}\|^2$ is a rank-one projector, and hence,

$$\log I\!\!E \exp\bigl(\xi \boldsymbol{u}\boldsymbol{u}^\top\bigr) = \log\bigl(I\!\!E \mathrm{e}^{\xi\|\boldsymbol{u}\|^2}\bigr) \frac{\boldsymbol{u}\boldsymbol{u}^\top}{\|\boldsymbol{u}\|^2}.$$

With $\boldsymbol{U}_i \stackrel{\mathrm{def}}{=} \boldsymbol{u}_i\boldsymbol{u}_i^\top / \|\boldsymbol{u}_i\|^2$ and $\xi_i = \theta(\varepsilon_i^2 - 1)$, we derive

$$\begin{aligned}
\log I\!\!E \exp\bigl\{\theta(\varepsilon_i^2 - 1)\boldsymbol{u}_i\boldsymbol{u}_i^\top\bigr\} &= \log I\!\!E \exp\bigl\{\theta(\varepsilon_i^2 - 1)\|\boldsymbol{u}_i\|^2\bigr\}\boldsymbol{U}_i \\
&= \log\left(\frac{\exp\bigl(-\|\boldsymbol{u}_i\|^2\theta\bigr)}{\sqrt{1 - 2\|\boldsymbol{u}_i\|^2\theta}}\right)\boldsymbol{U}_i \\
&= \Bigl\{-\|\boldsymbol{u}_i\|^2\theta - \frac{1}{2}\log(1 - 2\theta\|\boldsymbol{u}_i\|^2)\Bigr\}\boldsymbol{U}_i
\end{aligned}$$

and by (**??**)

$$\begin{aligned}
&I\!\!P\bigl(\|\boldsymbol{Z}\|_{\mathrm{op}} \ge z\bigr) \\
&\quad\le 2\inf_{\theta>0} \mathrm{e}^{-\theta z}\operatorname{tr}\exp\Bigl\{\sum_{i=1}^n \frac{\boldsymbol{u}_i\boldsymbol{u}_i^\top}{\|\boldsymbol{u}_i\|^2}\Bigl\{-\|\boldsymbol{u}_i\|^2\theta - \frac{1}{2}\log(1 - 2\theta\|\boldsymbol{u}_i\|^2)\Bigr\}\Bigr\}.
\end{aligned} \tag{13}$$

Denote $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)^\top$, where

$$\eta_i = -\theta - \frac{\log(1 - 2\|\boldsymbol{u}_i\|^2\theta)}{2\|\boldsymbol{u}_i\|^2}.$$

The use of (**??**) and (**??**) yields for $\theta < (2\delta^2)^{-1}$

$$\begin{aligned}
\eta_i &= \frac{1}{2\|\boldsymbol{u}_i\|^2}\bigl\{2\theta\|\boldsymbol{u}_i\|^2 - \log(1 - 2\theta\|\boldsymbol{u}_i\|^2)\bigr\} \\
&\le \frac{\bigl(2\theta\|\boldsymbol{u}_i\|^2\bigr)^2}{4\|\boldsymbol{u}_i\|^2(1 - 2\theta\delta^2)} \le \frac{\theta^2\delta^2}{(1 - 2\theta\delta^2)}.
\end{aligned}$$

Then by (**??**) and $\mathcal{U}\mathcal{U}^\top = I_p$ using $\mu = 2\theta\delta^2$

$$\begin{aligned}
I\!\!P\bigl(\|\boldsymbol{Z}\|_{\mathrm{op}} \ge t\bigr) &\le 2\inf_{\theta>0} \mathrm{e}^{-\theta z}\operatorname{tr}\exp\bigl\{\mathcal{U}\operatorname{diag}(\boldsymbol{\eta})\mathcal{U}^\top\bigr\} \le 2\inf_{\theta>0} \mathrm{e}^{-\theta z}\operatorname{tr}\exp\bigl\{\|\boldsymbol{\eta}\|_\infty I_p\bigr\} \\
&\le 2p\inf_{\theta>0}\exp\Bigl\{-\theta z + \frac{\theta^2\delta^2}{1 - 2\theta\delta^2}\Bigr\} = 2p\inf_{\mu>0}\exp\Bigl\{-\mu\frac{t}{2\delta^2} + \frac{\mu^2\delta^{-2}}{1-\mu}\Bigr\}.
\end{aligned}$$

Lemma **??** helps to bound for $\mathtt{x}_p = \mathtt{x} + \log(2p)$ and $t = 2\delta\mathtt{x}_p^{1/2} + 2\delta^2\mathtt{x}_p$ that

$$\inf_{\mu>0}\exp\Bigl\{-\mu\frac{t}{2\delta^2} + \frac{\mu^2\delta^{-2}}{1-\mu}\Bigr\} = \inf_{\mu>0}\Bigl\{-\mu\bigl(\delta^{-1}\mathtt{x}_p^{1/2} + \mathtt{x}_p\bigr) + \frac{\mu^2\delta^{-2}}{4(1-\mu)}\Bigr\} \le -\mathtt{x}_p.$$

Therefore,

$$I\!\!P\left(\|\boldsymbol{Z}\|_{\mathrm{op}} \ge 2\delta\sqrt{\mathtt{x}_p} + 2\delta^2\mathtt{x}_p\right) \le 2p\,\mathrm{e}^{-\mathtt{x}_p} = \mathrm{e}^{-\mathtt{x}}$$

as required. $\square$

Consequence. For non i.i.d $\varepsilon = R^{1/2}\xi$

$$\mathbb{P}\Bigl(\|\boldsymbol{Z}\|_{\mathrm{op}} \ge \lambda_{\max}(R)(2\delta\sqrt{\mathtt{x}_p} + 2\delta^2\mathtt{x}_p)\Bigr) \le \mathrm{e}^{-\mathtt{x}}$$

Consequence.

$$\mathbb{P}\Bigl(\|\mathcal{U}\operatorname{diag}(\eta)\mathcal{U}^T\|_{\mathrm{op}} > z\Bigr) \le 2p\inf_{\mu}\exp\left\{-\mu z + \max_i \frac{1}{\|u_i\|^2}\log \mathbb{E}e^{\mu\|u_i\|^2\eta_i}\right\}$$

Theorem 7 (Deviation bound for matrix Gaussian sums)**.** Let vectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$ in $\mathbb{R}^p$ satisfy

$$\|\boldsymbol{u}_i\| \le \delta$$

for a fixed constant δ. Let ε_i be independent standard Gaussian, $i = 1, \ldots, n$. Then for each vector $\boldsymbol{B} = (b_1, \ldots, b_n)^\top \in \mathbb{R}^n$, the matrix $\boldsymbol{Z}_1$ with

$$\boldsymbol{Z}_1 \stackrel{\mathrm{def}}{=} \sum_{i=1}^n \varepsilon_i b_i \boldsymbol{u}_i \boldsymbol{u}_i^\top$$

fulfills

$$\mathbb{P}\Bigl(\|\boldsymbol{Z}_1\|_{\mathrm{op}} \ge \delta^2\|\boldsymbol{B}\|\sqrt{2\mathtt{x}}\Bigr) \le 2\mathrm{e}^{-\mathtt{x}}.$$

Proof. As ε_i are i.i.d. standard normal and $\mathbb{E}\mathrm{e}^{a\varepsilon_i} = \mathrm{e}^{a^2/2}$ for $|a| < 1/2$, it follows from the Master inequality and Lemma 0.3.

$$\begin{aligned}
\mathbb{P}\bigl(\|\boldsymbol{Z}_1\|_{\mathrm{op}} \ge z\bigr) &\le 2\inf_{\theta>0}\mathrm{e}^{-\theta z}\operatorname{tr}\exp\Bigl\{\sum_{i=1}^n \log\mathbb{E}\exp(\theta\varepsilon_i b_i \boldsymbol{u}_i\boldsymbol{u}_i^\top)\Bigr\} \\
&\le 2\inf_{\theta>0}\mathrm{e}^{-\theta z}\operatorname{tr}\exp\Bigl\{\sum_{i=1}^n \frac{\theta^2 b_i^2\|\boldsymbol{u}_i\|^4}{2}\,\frac{\boldsymbol{u}_i\boldsymbol{u}_i^\top}{\|\boldsymbol{u}_i\|^2}\Bigr\}.
\end{aligned}$$

Moreover, as $\|\boldsymbol{u}_i\| \le \delta$ and $\boldsymbol{U}_i = \boldsymbol{u}_i\boldsymbol{u}_i^\top/\|\boldsymbol{u}_i\|^2$ is a rank-one projector with $\operatorname{tr}\boldsymbol{U}_i = 1$, it holds

$$\operatorname{tr}\exp\Bigl\{\frac{\theta^2}{2}\sum_{i=1}^n b_i^2\|\boldsymbol{u}_i\|^4\boldsymbol{U}_i\Bigr\} \le \exp\operatorname{tr}\Bigl(\frac{\theta^2\delta^4}{2}\sum_{i=1}^n b_i^2\boldsymbol{U}_i\Bigr) = \exp\frac{\theta^2\delta^4\|\boldsymbol{B}\|^2}{2}.$$

This implies for $z = \delta^2\|\boldsymbol{B}\|\sqrt{2\mathtt{x}}$

$$\mathbb{P}\bigl(\|\boldsymbol{Z}_1\|_{\mathrm{op}} \ge z\bigr) \le 2\inf_{\theta>0}\exp\Bigl(-\theta z + \frac{1}{2}\theta^2\delta^4\|\boldsymbol{B}\|^2\Bigr) = 2\mathrm{e}^{-\mathtt{x}}$$

and the assertion follows. □

Consequence. For non i.i.d $\varepsilon = R^{1/2}\xi$

$$\mathbb{P}\Bigl(\|\boldsymbol{Z}_1\|_{\mathrm{op}} \ge \sqrt{\lambda_{\max}(R)}\delta^2\|\boldsymbol{B}\|\sqrt{2\mathtt{x}}\Bigr) \le 2\mathrm{e}^{-\mathtt{x}}.$$