DRAFT: Application of Matrix Concentration Inequalities

Shend Thaqi

December 29, 2023

Contents

1 Concentration for an Operator Sum

Here, I will write why concentration inequalities for one-dimensional random variables do not generalize easily. The reason is that in the one-dimensional case, for a sum of independent random variables $X_1, \dots, X_n$ we can use the Markov inequality with the exponential function to obtain

$$\mathbb{E}\left[e^{\sum_{k=1}^n X_k}\right] = \prod_{k=1}^n \mathbb{E}\left[e^{X_k}\right].$$

This equality only works if we have commutativity of $X_1, \dots, X_n$. For non-commutative groups, we need to introduce some technical results. The approach in order to generalize such claims as shown in [Tro11b], [Tro15], [HNWTW20] suggests that we need to find some rather advanced result from functional analysis to overcome the problem of non-commutativity. With this result the usual concentration inequalities are shown in a similar way as in the commutative setting.

1.1 Lieb's theorem

An important tool in the one-dimensional case for sums of random variables is Markov's inequality. As it is important for sums of operators as well, we will present it as well.

Theorem 1.1 (Markov's inequality)**.**
Let X be a real-valued random variable, $\phi : \mathbb{R} \to [0, \infty)$ be increasing and $t > 0$. Assume $\phi(t) > 0$. Then,

$$\mathbb{P}(X \geq t) \leq \frac{1}{\phi(t)}\mathbb{E}[\phi(X)].$$

Proof. Since ϕ is increasing, $\{X \geq t\} \subseteq \{\phi(X) \geq \phi(t)\}$. Thus, $\mathbb{P}(X \geq t) \leq \mathbb{P}(\phi(X) \geq \phi(t))$. We also have

$$\mathbb{E}[\phi(X)] \geq \mathbb{E}[\phi(X)\mathbb{1}[X \geq t]] \geq \mathbb{E}[\phi(t)\mathbb{1}[X \geq t]] = \phi(t)\mathbb{P}(X \geq t).$$

□

Theorem 1.2 (Markov's inequality for traces)**.**
Let $\boldsymbol{X}$ be a random self-adjoint operator on $\mathcal{S}_1(\mathcal{H})$, let $\psi : \mathbb{R} \to [0, \infty)$ be an increasing function. Then, for $t > 0$

$$\mathbb{P}(\lambda_{\max}(\boldsymbol{X}) \geq t) \leq \frac{1}{\psi(t)}\mathbb{E}[\operatorname{tr}\psi(\boldsymbol{X})].$$

Proof. Since $\boldsymbol{X}$ is self-adjoint, $\phi(\boldsymbol{X}) \geq 0$ is well-defined. By spectral mapping and Lidskii's theorem

$$\phi(\lambda_{\max}(\boldsymbol{X})) = \lambda_{\max}(\phi(\boldsymbol{X})) \leq \sum_{\lambda \in \sigma_{\mathrm{p}}(\phi(X))} \lambda = \operatorname{tr}\phi(X).$$

It follows

$$\mathbb{P}(\lambda_{\max}(\boldsymbol{X}) \geq t) \leq \mathbb{P}(\lambda_{\max}(\phi(\boldsymbol{X})) \geq \phi(t)) \leq \mathbb{P}(\operatorname{tr}\phi(\boldsymbol{X}) \geq \phi(t)) \leq \frac{1}{\phi(t)}\mathbb{E}[\operatorname{tr}\phi(X)].$$

□

Lemma 1.3 (Finiteness of trace exponential)**.**
Let $\boldsymbol{S} \in \mathcal{S}_1(\mathcal{H})$. Then,

$$\operatorname{tr}\left[|\exp(\boldsymbol{S}) - 1|\right] < \infty.$$

Proof. Notice that for $x \geq 0$, we have

$$\begin{aligned} e^x - 1 &= \sum_{n=1}^\infty \frac{1}{n!}x^n \\ &\leq \sum_{n=0}^\infty \frac{1}{n!}x^{n+1} \\ &= x \cdot e^x \end{aligned}$$

Therefore,

$$\begin{aligned}\operatorname{tr}\left[|\exp(\boldsymbol{S})-1|\right] &= \sum_{k=1}^{\infty}\left|e^{\lambda_k}-1\right| \\ &\leq \sum_{k=1}^{\infty}|\lambda_k|\cdot e^{|\lambda_k|} \\ &\leq \operatorname{tr}|\boldsymbol{S}|\cdot e^{\|\boldsymbol{S}\|}.\end{aligned}$$

using spectral theory. □

Theorem 1.4 (Golden-Thompson, [Gol65], [Tho65])**.**
Let $\boldsymbol{S}, \boldsymbol{T}$ be self-adjoint operators such that $e^{\boldsymbol{S}}, e^{\boldsymbol{T}} \in \mathcal{S}_1(\mathcal{H})$. Then,

$$\operatorname{tr}\exp(\boldsymbol{S}+\boldsymbol{T}) \leq \operatorname{tr}(\exp(\boldsymbol{S})\cdot\exp(\boldsymbol{S})).$$

Theorem 1.5 (Lieb's Theorem, [Lie73])**.**
Let $\boldsymbol{T}$ be a self-adjoint operator such that $e^{\boldsymbol{T}} \in \mathcal{S}_1$. Then, the function

$$\left\{\boldsymbol{S}\in\mathcal{S}_1(\mathcal{H})\,\middle|\,e^{\boldsymbol{S}}\in\mathcal{S}_1(\mathcal{H})\right\}\to\mathbb{R},\ \boldsymbol{S}\mapsto\operatorname{tr}\exp(\boldsymbol{S}+\ln(\boldsymbol{T}))$$

is concave.

In the following when we refer to Lieb's inequality we will not write down the assumption that $e^T \in \mathcal{S}_1(\mathcal{H})$ for $T \in \mathcal{S}_1(\mathcal{H})$. If it does not hold, both sides of the inequality are infinite, and the result holds regardless.

Corollary 1.6 (Lieb's Theorem for random matrices)**.**
Let $\boldsymbol{S}$ be a self-adjoint trace-class operator, and let $\boldsymbol{X}$ be a random self-adjoint trace-class operator. Then,

$$\mathbb{E}[\operatorname{tr}\exp(\boldsymbol{S}+\boldsymbol{X})] \leq \operatorname{tr}\exp(\boldsymbol{S}+\ln\mathbb{E}[\exp(\boldsymbol{X})]).$$

1.2 Self-Adjoint Dilation

In this chapter, we will derive concentration results for the maximum eigenvalues of sums of random self-adjoint operators. In many applications, one might be interested to derive results for random operators that are not self-adjoint. In that case concentration results about the operator norm can be useful as the eigenvalues are not necessarily real. A technique called *self-adjoint dilation* can be employed to use concentration for self-adjoint operators to derive these without having to find new proofs.

Definition 1.1 (Self-adjoint dilation)**.**
Let $\boldsymbol{S} \in \mathcal{L}(\mathcal{H})$. We define the *self-adjoint dilation* of $\boldsymbol{S}$ as

$$\mathfrak{S}(\boldsymbol{S}) = \begin{bmatrix}\boldsymbol{0} & \boldsymbol{S}\\ \boldsymbol{S}^* & \boldsymbol{0}\end{bmatrix}.$$

If we consider $\mathfrak{S}(\boldsymbol{S})$ as an operator on $\mathcal{L}(\mathcal{H}^2)$, it can be shown readily that $\mathfrak{S}(\boldsymbol{S})$ is self-adjoint considering the scalar product $\langle\cdot\mid\cdot\rangle_{\mathcal{H}^2}$ defined by the relation

$$\left\langle\begin{bmatrix}\boldsymbol{u}\\ \boldsymbol{v}\end{bmatrix}\,\middle|\,\begin{bmatrix}\boldsymbol{w}\\ \boldsymbol{x}\end{bmatrix}\right\rangle_{\mathcal{H}^2} = \langle\boldsymbol{u}\mid\boldsymbol{w}\rangle+\langle\boldsymbol{v}\mid\boldsymbol{x}\rangle$$

1.3 Intrinsic Dimension Lemma

Here, I will write why the intrinsic dimension is useful to provide results for matrices of large dimension or even compact operators on separable possibly infinite-dimensional Hilbert spaces.

Definition 1.2 (Intrinsic dimension)**.**
Let $\boldsymbol{S}$ be an operator in $\mathcal{S}_1(\mathcal{H})$. Then, we define the *intrinsic dimension* of $\boldsymbol{S}$ as $\operatorname{int}\dim(\boldsymbol{S}) = \frac{\operatorname{tr}|\boldsymbol{S}|}{\|\boldsymbol{S}\|}$.

Notice that for a self-adjoint matrix $\boldsymbol{S}$ of size $d \times d$, we have that the maximum of the modulus of eigenvalues is equal to the largest singular value. Since the trace is the sum of the modulus of singular values, we have that

$$\operatorname{int}\dim(\boldsymbol{S}) = \frac{\operatorname{tr}|\boldsymbol{S}|}{\|\boldsymbol{S}\|} \leq d.$$

The intrinsic dimension can be interpreted as a metric for the point spectrum: The higher the intrinsic value is the closer each of the modulus of the eigenvalues are to the maximum of this set. In that sense it is not surprising that the intrinsic dimension offers a way to circumvent the dimension of the Hilbert space. This is useful to provide results that generalize for more general settings. However, it may be difficult to provide information about the intrinsic dimension of the bound for the random operator if the Hilbert space is infinite-dimensional.

Lemma 1.7 (Intrinsic dimension lemma, [Tro15]).
Let $\psi : [0, \infty) \to [0, \infty)$ be a convex function such that $\psi(0) = 0$. Let $\boldsymbol{A}$ be a positive operator. Then,

$$\operatorname{tr}(\psi(\boldsymbol{A})) \leq \operatorname{int}\dim(\boldsymbol{A}) \cdot \psi(\boldsymbol{A}).$$

Proof. The result is shown in [Tro15] for matrices. The same proof holds for trace-class, hence compact, operators as well using an approximation result. □

1.3.1 Intrinsic Chernoff Inequality

We first present the result for matrices as in [Tro15]. Then, we use an approximation argument to achieve the result for compact operators.

Theorem 1.8 (Matrix Chernoff lemma).
Let $\mathbf{0} \leq \boldsymbol{X} \leq \mathbf{1}$ for some random, self-adjoint matrix. Then, for $n \in \mathbb{N}$ and $\theta > 0$

$$\mathbb{E}\left[\exp(n\theta \boldsymbol{X}) - 1\right] \leq \frac{e^{n\theta} - 1}{n} \mathbb{E}[\boldsymbol{X}]$$

Proof. The result follows from spectral theory and the inequalities for real-valued functions. □

This lemma is used by [Tro15] to show the following theorem for matrices.

Theorem 1.9 (Intrinsic Matrix Chernoff, [Tro15]).
Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be independent, random, self-adjoint matrices of size $d \times d$ such that

$$\mathbf{0} \preceq \boldsymbol{X}_k \preceq \mathbf{1} \quad \text{almost surely for all } k \leq n.$$

Assume there is a matrix $\boldsymbol{M} \geq \mathbb{E}\left[\sum_{k=1}^{n} \boldsymbol{X}_k\right]$ and define $\mu_{\max} = \lambda_{\max}(\boldsymbol{M})$. Then,

$$\mathbb{P}\left(\lambda_{\max}\left(\sum_{k=1}^{n} \boldsymbol{X}_k\right) \geq (1+\epsilon)\mu_{\max}\right) \leq 2 \operatorname{int}\dim(\boldsymbol{M}) \cdot \left(\frac{e^{\epsilon}}{(1+\epsilon)^{1+\epsilon}}\right)^{\mu_{\max}} \qquad \text{for } \epsilon \geq 1/\mu_{\max}.$$

Since $\boldsymbol{X}$ does not appear evaluated at an nonlinear function it is straightforward to generalize this result for compact operators on a separable Hilbert space.

Corollary 1.10 (Intrinsic Operator Chernoff).
Let $\mathcal{H}$ be a separable Hilbert space. Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be independent, random, self-adjoint operators on $\mathcal{S}_1(\mathcal{H})$. Assume that

1) $\mathbf{0} \preceq \boldsymbol{X}_k \preceq \mathbf{1}$ almost surely;

2) there is a self-adjoint trace-class operator $\boldsymbol{M} \succeq \mathbb{E}\left[\sum_{k=1}^{n} X_k\right]$.

Then, for $\epsilon \geq 1/\mu_{\max}$

$$\mathbb{P}\left(\lambda_{\max}\left(\sum_{k=1}^{n} \boldsymbol{X}_k\right) \geq (1+\epsilon)\mu_{\max}\right) \leq 2 \operatorname{int}\dim(\boldsymbol{M}) \cdot \left(\frac{e^{\epsilon}}{(1+\epsilon)^{1+\epsilon}}\right)^{\mu_{\max}}.$$

Proof. $\boldsymbol{M}$ is positive and $\boldsymbol{X}_k^N$ is positive almost surely, therefore we have that $\boldsymbol{M}$ is a boundary for $\mathbb{E}\left[\sum_{k=1}^n \boldsymbol{X}_k^N\right]$ as well. By construction, we have $\lim_{N\to\infty} \operatorname{tr}\left|\boldsymbol{X}^N - \boldsymbol{X}\right| = 0$. Therefore, we have $\lim_{N\to\infty} \lambda_{\max}\left(\boldsymbol{X}^N\right) = \lambda_{\max}\left(\boldsymbol{X}\right)$. We also have $\lambda_{\max}(\boldsymbol{X}^N) \leq \lambda_{\max}\left(\boldsymbol{X}\right)$ since $\boldsymbol{X}^N = \boldsymbol{P}_N \boldsymbol{X} \boldsymbol{P}_N$. Thus,

$$\{\lambda_{\max}(\boldsymbol{X}^N) > t\} \nearrow \{\lambda_{\max}\left(\boldsymbol{X}\right) > t\}.$$

Therefore,

$$\lim_{N\to\infty} \mathbb{P}\left(\lambda_{\max}\left(\sum_{k=1}^n \boldsymbol{X}^N\right) \leq t\right) = \mathbb{P}\left(\lambda_{\max}\left(\sum_{k=1}^n \boldsymbol{X}_k^N\right) \geq t\right)$$

□

The intrinsic matrix Chernoff concentation inequality was shown by [Tro15]. They mention the curiosity that their approach does not work for the minimum eigenvalue. In the one-dimensional case there is such a lower bound. Considering the situation in an infinite-dimensional Hilbert space, this becomes clear. Compact operators attain an eigenvalue at zero or they have an accumulation point of eigenvalues at zero. As a result there cannot exist such a lower bound for the minimum eigenvalue. In the case of an infinite-dimensional Hilbert space, we also cannot consider $\sum_{k=1}^n (\boldsymbol{X}_k - \mathbf{1})$. While $\boldsymbol{X}_k - \mathbf{1}$ satisfies independence, self-adjointness and is bounded by $\mathbf{0}$ and $\mathbf{1}$, it is not a trace-class operator and thus not bounded by one.

1.3.2 Intrinsic Bernstein Inequality: Bounded Case

Here I will write about: Utility of the Bernstein inequality. What are its advantages: If the variance of the random operators is low it offers better bounds for the tail probability.

Theorem 1.11 (Intrinsic operator Bernstein inequality for bounded operators)**.**
Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be random, independent, centered operators. Assume there exists $L > 0$ such that

$$\|\boldsymbol{X}_k\| \leq \boldsymbol{L} \quad \text{almost surely for all } k \leq n.$$

Furthermore, let $\boldsymbol{V}_1^2, \boldsymbol{V}_2^2 \in \mathcal{S}_1(\mathcal{H})$ satisfy

$$\boldsymbol{V}_1^2 \geq \sum_{k=1}^n \mathbb{E}\left[\boldsymbol{X}_k \boldsymbol{X}_k^*\right]$$
$$\boldsymbol{V}_2^2 \geq \sum_{k=1}^n \mathbb{E}\left[\boldsymbol{X}_k^* \boldsymbol{X}_k\right].$$

Define

$$d = \operatorname{int}\dim \begin{bmatrix} \boldsymbol{V}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{V}_2 \end{bmatrix} \quad \text{and} \quad v = \max\{\|\boldsymbol{V}_1\|, \|\boldsymbol{V}_2\|\}.$$

Then,

$$\mathbb{P}\left(\left\|\sum_{k=1}^n \boldsymbol{X}_k\right\| \geq t\right) \leq 4d \cdot \exp\left(\frac{-t^2/2}{v + Lt/3}\right).$$

This result has been proven for finite-dimensional matrices in [Tro15]. We want to give the generalization for trace-class operators.

Proof. The approach to the proof is identical to the generalization of the matrix Chernoff inequality. □

1.3.3 Intrinsic Hoeffding Inequality: Sub-Gaussian Case

Here I will write about: The Hoeffding inequality for sub-Gaussians is based on the Hoeffding lemma which caracterizes sub-Gaussian random variables and gives a bound for the moment generating function. In the matrix case, we have to introduce a similar lemma. (!!! What is even the definition for sub-Gaussian matrices which resembles the original definition of sub-Gaussian random variables).

We first establish the Hoeffding inequality for a sum of sub-Gaussian matrices with the intrinsic dimension. Then, we use an approximation argument to generalize the result for compact operators.

Theorem 1.12 (Intrinsic operator Hoeffding: Sub-Gaussian Case)**.**
Let $X_1, \dots, X_n$ be random, centered, independent, self-adjoint operators almost surely in $\mathcal{S}_1(\mathcal{H})$ such that

$$\mathbb{E}[\exp(\theta X_k)] \leq \exp(\frac{1}{2}\theta^2 V_k^2)$$

for some fixed positive operators $V_k^2 \in \mathcal{S}_1(\mathcal{H})$ and all $\theta \in \mathbb{R}$. Then, for $t > 0$

$$\mathbb{P}(\lambda_{\max}(\sum_k X_k) \geq t) \leq 2 \operatorname{int\,dim}\left(\sum_{k=1}^n V_k^2\right) \exp\left(-\frac{t^2}{2\|\sum_{k=1}^n V_k^2\|}\right).$$

Proof. By Markov inequality, we have

$$\mathbb{P}(\lambda_{\max}(\sum_{k=1}^n X_k) \geq t) \leq \frac{1}{\phi(t)}\mathbb{E}\left[\operatorname{tr}\left(\phi(\sum_{k=1}^n X_k)\right)\right]$$

for $\phi(t) = e^{\theta t} - 1$. To apply Lieb's theorem, notice that

$$\begin{aligned}
\mathbb{E}\left[\operatorname{tr}\left(\exp\left(\theta\sum_{k=1}^n X_k\right) - 1\right)\right] &= \mathbb{E}\left[\mathbb{E}\left[\operatorname{tr}\left(\exp\left(\theta\sum_{k=1}^{n-1} X_k + \theta X_n\right) - 1)\right)\middle|\mathcal{F}_{n-1}\right]\right] \\
&\leq \mathbb{E}\left[\operatorname{tr}\left(\exp\left(\theta\sum_{k=1}^n X_k + \ln\mathbb{E}\left[\exp\left(\theta X_k\right)\right]\right) - 1\right)\right] \\
&\leq \dots \\
&\leq \mathbb{E}\left[\operatorname{tr}\left(\exp\left(\sum_{k=1}^n \ln\left(\mathbb{E}\left[\exp\left(\theta X_k\right)\right)\right]\right) - 1\right)\right] \\
&\leq \operatorname{tr}\left(\exp\left(\frac{1}{2}\theta^2\sum_{k=1}^n V_k^2\right) - 1\right).
\end{aligned}$$

ϕ is convex and satisfies $\phi(0) = 0$. Notice that the last term is finite by Lemma 1.3. Therefore, the intrinsic dimension lemma is applicable and we obtain

$$\mathbb{E}\left[\operatorname{tr}\left(\exp\left(\frac{1}{2}\theta^2\sum_{k=1}^n V_k^2\right) - 1\right)\right] \leq \operatorname{int\,dim}\left(\sum_{k=1}^n V_k^2\right) \cdot e^{\frac{1}{2}\theta^2\|\sum_{k=1}^n V_k^2\|}$$

Using our results earlier, we get

$$\mathbb{P}(\lambda_{\max}(\sum_{k=1}^n X_k) \geq t) \leq \operatorname{int\,dim}\left(\sum_{k=1}^n V_k^2\right) \cdot \frac{1}{e^{\theta t} - 1} e^{\frac{1}{2}\theta^2\|\sum_{k=1}^n V_k^2\|}$$

Notice that

$$\frac{1}{e^{\theta t} - 1} e^{\frac{1}{2}\theta^2\|\sum_{k=1}^n V_k^2\|} = \frac{e^{\theta t}}{e^{\theta t} - 1} e^{-\theta t + \frac{1}{2}\theta^2\|\sum_{k=1}^n V_k^2\|}$$

We have $\frac{e^{\theta t}}{e^{\theta t}-1} = 1 + \frac{1}{e^{\theta t}-1} \leq 1 + \frac{1}{\theta t}$ as $1 + x \leq e^x$ for $x \in \mathbb{R}$. We also have that $\psi(\theta) = -\theta t + \frac{1}{2}\theta^2\|\sum_{k=1}^n V_k^2\|$ is minimized at $\theta^* = \frac{t}{\|\sum_{k=1}^n V_k^2\|}$ and $\psi(\theta^*) = -\frac{t^2}{2\|\sum_{k=1}^n V_k^2\|}$. Thus,

$$\mathbb{P}(\lambda_{\max}(\sum_{k=1}^n X_k) \geq t) \leq \operatorname{int\,dim}\left(\sum_{k=1}^n V_k^2\right)\left(1 + \frac{\|\sum_{k=1}^n V_k^2\|}{t^2}\right) e^{-\frac{t^2}{2\|\sum_{k=1}^n V_k^2\|}}$$

If $t \geq \|\sum_{k=1}^n V_k^2\|$, we have $\left(1 + \frac{\|\sum_{k=1}^n V_k^2\|}{t^2}\right) \leq 2$. If $t^2 < \|\sum_{k=1}^n V_k^2\|$, we have that

$$\left(1 + \frac{\|\sum_{k=1}^n V_k^2\|}{t^2}\right) e^{-\frac{t^2}{2\|\sum_{k=1}^n V_k^2\|}} \geq 2e^{-\frac{1}{2}} \geq 1$$

In this case we have that

$$\mathbb{P}(\lambda_{\max}(\sum_{k=1}^{n} X_k) \geq t) \leq 1 \leq 2\,\mathrm{int\,dim}\left(\sum_{k=1}^{n} V_k^2\right) e^{-\frac{t^2}{2\left\|\sum_{k=1}^{n} V_k^2\right\|}}$$

Thus in both cases we achieve the desired result. □

Notice that using a boundary $\sum_{k=1}^{n} V_k^2$ instead of $\sum_{k=1}^{n} V_k$ is purely symbolical and is related to the variance of the sum $\sum_{k=1}^{n} X_k$.

1.3.4 Intrinsic Bernstein Inequality: Sub-Exponential Case

Similarly to the sum of sub-Gaussian concentration result, we first state the result for matrices and then generalize the assertion with a similar approximation result. More elaboration (!!!: Similarly to sub-Gaussian case, can we find a matrix equivalence lemma as in real-value case?)

Theorem 1.13 (Intrinsic operator Bernstein: sub-Gaussian case)**.**
Let $X_1, \dots, X_n$ be random, centered, independent, self-adjoint operators almost surely in $\mathcal{S}_1(\mathcal{H})$ such that for some fixed self-adjoint matrices $V_k \in \mathcal{S}_1(\mathcal{H})$ and all $|\theta| \leq \frac{1}{b_k}$

$$\mathbb{E}[\exp(\theta X_k)] \leq \exp(\frac{1}{2}\theta^2 V_k^2)$$

. Then, for $a_1, \dots, a_n \in \mathbb{R}$ and $t > 0$

$$\mathbb{P}(\lambda_{\max}(\sum_{k=1}^{n} a_k X_k) \geq t) \leq 2\,\mathrm{int\,dim}\left(\sum_{k=1}^{n} a_k^2 V_k^2\right) \exp\left(-\min\left\{\frac{t^2}{2\|\sum_{k=1}^{n} a_k^2 V_k^2\|}, \frac{t}{2\max_k |a_k b_k|}\right\}\right).$$

Proof. Assume without loss of generality that $a_k > 0$. By Markov inequality, we have

$$\mathbb{P}(\lambda_{\max}(\sum_{k=1}^{n} a_k X_k) \geq t) \leq \frac{1}{\phi(t)} \mathbb{E}\left[\mathrm{tr}\left(\phi(\sum_{k=1}^{n} a_k X_k)\right)\right]$$

for $\phi(t) = e^{\theta t} - 1$ where $|\theta| \leq \frac{1}{\max_k a_k b_k}$.

To apply Lieb's theorem, notice that for $|\theta| \leq \frac{1}{\max_k a_k b_k}$

$$\begin{aligned}
\mathbb{E}\left[\mathrm{tr}\left(\exp\left(\theta \sum_{k=1}^{n} a_k X_k\right) - 1\right)\right] &= \mathbb{E}\left[\mathbb{E}\left[\mathrm{tr}\left(\exp\left(\theta \sum_{k=1}^{n-1} a_k X_k + \theta a_n X_n\right) - 1)\right) \middle| \mathcal{F}_{n-1}\right]\right] \\
&\leq \mathbb{E}\left[\mathrm{tr}\left(\exp\left(\theta \sum_{k=1}^{n} a_k X_k + \ln \mathbb{E}\left[\exp\left(\theta a_k X_k\right)\right]\right) - 1\right)\right] \\
&\leq \dots \\
&\leq \mathbb{E}\left[\mathrm{tr}\left(\exp\left(\theta \sum_{k=1}^{n} a_k X_k \ln\left(\mathbb{E}\left[\exp\left(\theta a_k X_k\right)\right)\right]\right)\right) - 1\right)\right] \\
&\leq \mathrm{tr}\left(\exp\left(\frac{1}{2}\theta^2 \sum_{k=1}^{n} a_k^2 V_k^2\right) - 1\right).
\end{aligned}$$

ϕ is convex and satisfies $\phi(0) = 0$. Notice that the last term is finite by Lemma 1.3. Therefore, the intrinsic dimension lemma is applicable and we obtain for $|\theta| \leq \frac{1}{\max_k a_k b_k}$

$$\mathbb{E}\,\mathrm{tr}\left(\exp\left(\frac{1}{2}\theta^2 \sum_{k=1}^{n} a_k^2 V_k^2\right) - 1\right) \leq \mathrm{int\,dim}\left(\sum_{k=1}^{n} a_k^2 V_k^2\right) \cdot e^{\frac{1}{2}\theta^2 \left\|\sum_{k=1}^{n} a_k^2 V_k^2\right\|}$$

Using our results earlier, we get

$$\mathbb{P}(\lambda_{\max}(\sum_{k=1}^{n} a_k X_k) \geq t) \leq \mathrm{int\,dim}\left(\sum_{k=1}^{n} a_k^2 V_k^2\right) \cdot \frac{1}{e^{\theta t} - 1} e^{\frac{1}{2}\theta^2 \left\|\sum_{k=1}^{n} a_k^2 V_k^2\right\|}$$

Notice that

$$\frac{1}{e^{\theta t}-1}e^{\frac{1}{2}\theta^2\left\|\sum_{k=1}^n a_k^2 V_k^2\right\|} = \frac{e^{\theta t}}{e^{\theta t}-1}e^{-\theta t+\frac{1}{2}\theta^2\left\|\sum_{k=1}^n a_k^2 V_k^2\right\|}$$

We have $\frac{e^{\theta t}}{e^{\theta t}-1} = 1+\frac{1}{e^{\theta t}-1} \leq 1+\frac{1}{\theta t}$ as $1+x \leq e^x$ for $x \in \mathbb{R}$. We also have that $\psi(\theta) = -\theta t + \frac{1}{2}\theta^2\left\|\sum_{k=1}^n a_k^2 V_k^2\right\|$ is minimized at $\theta^* = \min\left\{\frac{t}{\left\|\sum_{k=1}^n a_k^2 V_k^2\right\|}, \frac{1}{\max_k a_k b_k}\right\}$. Thus,

$$\mathbb{P}(\lambda_{\max}(\sum_{k=1}^n a_k X_k) \geq t) \leq \operatorname{int}\dim\left(\sum_{k=1}^n a_k^2 V_k^2\right)\left(1+\frac{1}{\theta^* t}\right)e^{-\frac{\theta^* t}{2}}$$

If $t \geq \frac{1}{\theta^*}$, we have $\left(1+\frac{1}{\theta^* t}\right) \leq 2$. If $t < \frac{1}{\theta^*}$, we have that

$$\left(1+\frac{1}{\theta^* t}\right)e^{-\frac{\theta^* t}{2}} \geq 1$$

In this case we have that

$$\mathbb{P}(\lambda_{\max}(\sum_{k=1}^n a_k X_k) \geq t) \leq 1 \leq 2\operatorname{int}\dim\left(\sum_{k=1}^n a_k^2 V_k^2\right)e^{-\frac{\theta^* t}{2}}$$

Thus in both cases we achieve the desired result. □

(!!! should we shorten the proof using better notation $V_a^2 = \sum_k a_k^2 V_k^2$. The proof is almost identical to the sub-Gaussian case. Can we extract arguments and introduce them earlier as lemmas?

1.3.5 Intrinsic Azuma inequality

Here I will write about symmetrization. In the one-dimensional case, we have some lemma to symmetrize the random variables such that we can conduct the proof as shown in !!! Here I insert some literature.

Lemma 1.14 (Rademacher lemma)**.**
Let A be a self-adjoint operator, ϵ be a Rademacher random variable. Then,

$$\mathbb{E}\left[e^{\epsilon\theta A}\right] \leq e^{\frac{1}{2}\theta^2 A^2}$$

Proof. We have that for $x \in \mathbb{R}$, $\frac{1}{2}e^x + \frac{1}{2}e^{-x} \leq e^{\frac{1}{2}x^2}$. By spectral theory,

$$\mathbb{E}\left[e^{\epsilon A}\right] = \frac{1}{2}e^{-A} + \frac{1}{2}e^{A} \leq e^{\frac{1}{2}A^2}$$

□

Lemma 1.15 (Symmetrization)**.**
Let X be a random self-adjoint matrix, A be a fixed self-adjoint matrix. Assume $X^2 \leq A^2$. Let ϵ be an independent Rademacher variable. Then,

$$\ln\mathbb{E}\left[e^{2\epsilon\theta X} \,\middle|\, X\right] \leq 2\theta^2 A^2$$

Proof. We have

$$\mathbb{E}\left[e^{2\epsilon\theta X} \,\middle|\, X\right] \leq e^{2\theta^2 X^2}$$

□

Here, we do not need equivalent statements for operators, since we will use the matrix-version´of Azuma's inequality to deduce the operator-version.

Some text about "cost of symmetrization". Mention, that for symmetrically-distributed random matrices, we have sharper bounds.

Theorem 1.16 (Intrinsic operator Azuma)**.**
Let $X, 1 \ldots, X_n$ be random self-adjoint operators almost surely in $\mathcal{S}_1(\mathcal{H})$ such that

1) $\mathbb{E}[X_{k+1} \,|\, \mathcal{F}_k] = 0$;

2) $X_k^2 \le A_k^2$ almost surely for some positive operators in $A_k^2 \in \mathcal{S}_1(\mathcal{H})$. Define $\sigma^2 := \left\|\sum_{k=1}^n A_k^2\right\|$.

Then,

$$\mathbb{P}(\lambda_{\max}(\sum_{k=1}^n X_k) \ge t) \le \operatorname{int}\dim\left(\sum_{k=1}^n A_k^2\right) \cdot \exp\left(-\frac{t^2}{8\sigma^2}\right)$$

Proof. By Markov inequality, we have

$$\mathbb{P}(\lambda_{\max}(\sum_{k=1}^n X_k) \ge t) \le \frac{1}{\phi(t)}\mathbb{E}\left[\operatorname{tr}\left(\phi(\sum_{k=1}^n X_k)\right)\right]$$

for $\phi(t) = e^{\theta t} - 1$. To apply Lieb's theorem, notice that

$$\begin{aligned}
\mathbb{E}\left[\operatorname{tr}\left(\exp\left(\theta\sum_{k=1}^n X_k\right) - 1\right)\right] &= \mathbb{E}\left[\mathbb{E}\left[\operatorname{tr}\left(\exp\left(\theta\sum_{k=1}^{n-1} X_k + \theta X_n\right) - 1)\right) \,\middle|\, \mathcal{F}_{n-1}\right]\right] \\
&\le \mathbb{E}\left[\mathbb{E}\left[\operatorname{tr}\left(\exp\left(\theta\sum_{k=1}^{n-1} X_k + 2\epsilon\theta X_n\right) - 1)\right) \,\middle|\, \mathcal{F}_{n-1}\right]\right] \\
&\le \mathbb{E}\left[\operatorname{tr}\left(\exp\left(\theta\sum_{k=1}^{n} X_k + \ln\mathbb{E}\left[\exp\left(2\epsilon\theta X_k\right)\right]\right) - 1\right)\right] \\
&\le \dots \\
&\le \mathbb{E}\left[\operatorname{tr}\left(\exp\left(\sum_{k=1}^{n} \ln\left(\mathbb{E}\left[\exp\left(2\epsilon\theta X_k\right)\right)\right]\right) - 1\right)\right] \\
&\le \operatorname{tr}\left(\exp\left(2\theta^2\sum_{k=1}^n A_k^2\right) - 1\right)
\end{aligned}$$

ϕ is convex and satisfies $\phi(0) = 0$. Notice that the last term is finite by Lemma 1.3. Therefore, the intrinsic dimension lemma is applicable and we obtain

$$\mathbb{E}\left[\operatorname{tr}\left(\exp\left(2\theta^2\sum_{k=1}^n A_k^2\right) - 1\right)\right] \le \operatorname{int}\dim\left(\sum_{k=1}^n A_k^2\right) \cdot e^{2\theta^2\left\|\sum_{k=1}^n A_k^2\right\|}$$

Using our results earlier, we get

$$\mathbb{P}(\lambda_{\max}(\sum_{k=1}^n X_k) \ge t) \le \operatorname{int}\dim\left(\sum_{k=1}^n A_k^2\right) \cdot \frac{1}{e^{\theta t} - 1} e^{2\theta^2\left\|\sum_{k=1}^n A_k^2\right\|}$$

Notice that

$$\frac{1}{e^{\theta t} - 1} e^{2\theta^2\left\|\sum_{k=1}^n A_k^2\right\|} = \frac{e^{\theta t}}{e^{\theta t} - 1} e^{-\theta t + 2\theta^2\left\|\sum_{k=1}^n A_k^2\right\|}$$

We have $\frac{e^{\theta t}}{e^{\theta t}-1} = 1 + \frac{1}{e^{\theta t}-1} \le 1 + \frac{1}{\theta t}$ as $1 + x \le e^x$ for $x \in \mathbb{R}$. We also have that $\psi(\theta) = -\theta t + 2\theta^2\left\|\sum_{k=1}^n A_k^2\right\|$ is minimized at $\theta^* = \frac{t}{\left\|4\sum_{k=1}^n A_k^2\right\|}$ and $\psi(\theta^*) = -\frac{t^2}{8\left\|\sum_{k=1}^n A_k^2\right\|}$. Thus,

$$\mathbb{P}(\lambda_{\max}(\sum_{k=1}^n X_k) \ge t) \le \operatorname{int}\dim\left(\sum_{k=1}^n A_k^2\right)\left(1 + \frac{4\left\|\sum_{k=1}^n A_k^2\right\|}{t^2}\right) e^{-\frac{t^2}{8\left\|\sum_{k=1}^n A_k^2\right\|}}$$

If $t \ge 4\left\|\sum_{k=1}^n A_k^2\right\|$, we have $\left(1 + \frac{4\left\|\sum_{k=1}^n A_k^2\right\|}{t^2}\right) \le 2$. If $t^2 < 4\left\|\sum_{k=1}^n A_k^2\right\|$, we have that

$$\left(1 + \frac{4\left\|\sum_{k=1}^n A_k^2\right\|}{t^2}\right) e^{-\frac{t^2}{8\left\|\sum_{k=1}^n A_k^2\right\|}} \ge 2e^{-\frac{1}{2}} \ge 1$$

In this case we have that

$$\mathbb{P}(\lambda_{\max}(\sum_{k=1}^n X_k) \geq t) \leq 1 \leq 2\,\mathrm{int}\dim\left(\sum_{k=1}^n A_k^2\right)\exp\left(-\frac{t^2}{8\|\sum_{k=1}^n A_k^2\|}\right)$$

Thus in both cases we achieve the desired result. □

One of the most immediate applications from Azuma's inequality in the one-dimensional case is McDiarmid's inequality (!!! Reference), also known as the bounded differences inequality. A similar corollary holds for operators as well.

Corollary 1.17 (Intrinsic operator McDiarmid)**.**
Let $Z_1, \dots, Z_n$ be real-valued indepedenent random variables. Let

$$H : \mathbb{R}^n \to \mathcal{S}_1(\mathcal{H}) \cap \mathcal{L}_s(\mathcal{H})$$

be a map such that there exists A_k such that for all $z_1, \dots, z_n, z_k'$, we have

$$\left(H(z_1, \dots, z_k, \dots, z_n) - H(z_1, \dots, z_k', \dots, z_n)\right)^2 \leq A_k^2$$

Define $\sigma^2 := \sum_{k=1}^n A_k^2$. Then, for $t > 0$, we have

$$\mathbb{P}\left(\lambda_{\max}\left(H(z) - \mathbb{E}[H(z)]\right) \geq t\right) \leq \mathrm{int}\dim\left(\sum_{k=1}^n A_k^2\right) \cdot e^{-\frac{t^2}{8\sigma^2}}$$

where $z = (Z_1, \dots, Z_n)$.

Proof. We use the argument as in [Tro11b]. Their proof holds similarly as in the finite-dimensional case, only that we need the validity of Azuma's operator inequality. For the sake of completeness, we present the proof. Let $\mathbb{E}_Z$ denote the expectation with respect to Z, holding the other random variables fixed. Define

$$Y_k = \mathbb{E}\left[H(z) \,\middle|\, \underbrace{Z_1, \dots, Z_k}_{\sigma(\mathcal{F}_k)}\right] = \mathbb{E}_{Z_{k-1}} \dots \mathbb{E}_{Z_n}\left[H(z)\right]$$

and $X_k = Y_k - Y_{k-1}$. Then X_k satisfies the following:

1) X_k is self-adjoint random operator;

2) $\mathbb{E}[X_{k+1} \,|\, \mathcal{F}_k] = 0$;

3) $X_k^2 \leq A_k^2$.

Self-adjointness is clear, as the space of self-adjoint operators is linear. The martingale property is clear as

$$\mathbb{E}[\mathbb{E}[H(z) \,|\, \mathcal{F}_{k+1}] - \mathbb{E}[H(z) \,|\, \mathcal{F}_k] \,|\, \mathcal{F}_k] = 0$$

using the tower property of conditional expectations. We now show the bound for the differences: Let Z_k' be an independent copy of Z_k, define $z' = (Z_1, \dots, Z_k', \dots, Z_n)$. Then, $\mathbb{E}_{Z_k}\left[H(z)\right] = \mathbb{E}_{Z_{k'}}\mathbb{E}\left[H(z')\right]$. Then we have

- $X_k = \mathbb{E}_{Z_{k+1}}\mathbb{E}_{Z_n}\mathbb{E}_{Z_k'}\left[H(z) - H(z')\right]$ since $H(z)$ is independent of Z_k';
- $(H(z) - H(z'))^2 \leq A_k^2$ by assumption.

An application of Jensen's inequality yields

$$X_k^2 = \left(\mathbb{E}_{Z_{k+1}} \dots \mathbb{E}_{Z_n}\mathbb{E}_{Z_k'}\left[H(z) - H(z')\right]\right)^2 \leq \mathbb{E}_{Z_{k+1}} \dots \mathbb{E}_{Z_n}\mathbb{E}_{Z_k'}\mathbb{E}\left[\left(H(z) - H(z')\right)^2\right] \leq A_k^2$$

This allows us to use Azuma's inequality which yields the result. □

1.3.6 Intrinsic Freedman Inequality

Theorem 1.18 (Intrinsic Freedman inequality)**.**
Let $(X_k)_k$ be an adapted sequence of random self-adjoint operators on a separable Hilbert space $\mathcal{H}$ such that

1) $X_k \in \mathcal{S}_1(\mathcal{H})$ almost surely;
2) $\|X_k\| \leq 1$ almost surely;
3) $\mathbb{E}_{k-1}[X_k] = 0$.

Define $Y_k := \sum_{j=1}^k X_j$ and $W_k = \sum_{j=1}^k \mathbb{E}_j \left[X_j^2\right]$. Then, for $t, \sigma^2 > 0$,

$$\mathbb{P}\left(\exists k \geq 0 : \lambda_{\max}(Y_k) \geq t \text{ and } \|W_k\| \leq \sigma^2\right) 4 \cdot \exp\left(-\frac{\sigma^2}{R^2} \cdot h\left(\frac{R}{\sigma}\right)\right)$$

where $h(u) = (1+u)\ln(1+u) - u$ for $u \geq 0$.

Proof. To be proven. □

It remains unclear how to prove a version of Freedman's inequality similar to Theorem 1.1 in [Tro11a] including the intrinsic dimension. The technical difficulties lie in the following: For matrices of size $d \times d$

2 Concentration for Products

2.1 Simple Products

Theorem 2.1 (Concentration for products)**.**
Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be independent random matrices of size $d \times d$. Assume that

1) there is a matrix $\boldsymbol{X}$ of size $d \times d$ such that for all $k = 1, \ldots, n$, we have $\mathbb{E}[\boldsymbol{X}_k] = \boldsymbol{X}$;
2) there are $a, r, s > 0$ such that for all $k = 1, \ldots, n$, we have
   i) $\|\boldsymbol{X}_k\| \leq a$ almost surely;
   ii) $\|\boldsymbol{X}_k - \boldsymbol{X}\| \leq r$ almost surely;
   iii) $\mathbb{E}\|\boldsymbol{X}_k - \boldsymbol{X}\| \leq s$.

Then, for $t > 0$

$$\mathbb{P}\left(\left\|\prod_{k=1}^n \boldsymbol{X}_k - \boldsymbol{X}^n\right\| \geq t\right) \leq 2d \cdot \exp\left(\frac{-t^2/2}{a^{n-1}(na^{n-1}s^2 + rt/3)}\right).$$

Proof. Use the notation $\boldsymbol{X}^{(k)} = \boldsymbol{X}_k \cdots \boldsymbol{X}_1$ and $\mathbb{E}_k[\,\cdot\,] = \mathbb{E}[\,\cdot \mid \boldsymbol{X}_1, \ldots, \boldsymbol{X}_k]$. Then, by independence of the random matrices

$$\begin{aligned}\left\|\boldsymbol{X}^{(n)} - \boldsymbol{X}^n\right\| &= \left\|\boldsymbol{X}^{(n)} - \mathbb{E}\left[\boldsymbol{X}^{(n)}\right]\right\| \\ &= \left\|\sum_{k=1}^n \left(\mathbb{E}_k\left[\boldsymbol{X}^{(n)}\right] - \mathbb{E}_{k-1}\left[\boldsymbol{X}^{(n)}\right]\right)\right\|.\end{aligned}$$

Denote $\boldsymbol{M}_k = \mathbb{E}_k[\boldsymbol{X}^{(n)}]$. Then, $(\boldsymbol{M}_k)_k$ forms a martingale adapted to the filtration $(\mathfrak{F}_k)_k$ defined by $\mathfrak{F}_k = \sigma(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_k)$. To apply Lemma 2.3, notice that

1)

$$\begin{aligned}\boldsymbol{M}_k - \boldsymbol{M}_{k-1} &= \mathbb{E}\left[\boldsymbol{X}_n \cdots \boldsymbol{X}_{k+1}\right]\left(\boldsymbol{X}_k \cdots \boldsymbol{X}_1\right) - \mathbb{E}\left[\boldsymbol{X}_n \cdots \boldsymbol{X}_k\right]\left(\boldsymbol{X}_{k-1} \cdots \boldsymbol{X}_1\right) \\ &= \boldsymbol{X}^{n-k}\left(\boldsymbol{X}_k - \boldsymbol{X}\right)\left(\boldsymbol{X}^{(k-1)}\right).\end{aligned}$$

Therefore taking the norm on both sides and using its submultiplicativity yields

$$\begin{aligned}\|\boldsymbol{M}_k - \boldsymbol{M}_{k-1}\| &\leq \|\boldsymbol{X}\|^{n-k} \cdot \|\boldsymbol{X}_k - \boldsymbol{X}\| \cdot \left\|\boldsymbol{X}^{(k-1)}\right\| \\ &\leq a^{n-1} r \quad \text{almost surely.}\end{aligned}$$

Thus, we can define $R = a^{n-1} r$.

2) Notice that for $k = 1, \ldots, n$

$$\begin{aligned}&\mathbb{E}_{k-1}\left[\left(\boldsymbol{M}_k - \boldsymbol{M}_{k-1}\right)\left(\boldsymbol{M}_k - \boldsymbol{M}_{k-1}\right)^*\right] \\ &\quad = \mathbb{E}_{k-1}\left[\boldsymbol{X}^{n-1}\left(\boldsymbol{X}_k - \boldsymbol{X}\right)\boldsymbol{X}^{k-1} \cdot \left(\boldsymbol{X}^{k-1}\right)^*\left(\boldsymbol{X}^* - \boldsymbol{X}_k^*\right)\left(\boldsymbol{X}^*\right)^{n-k}\right].\end{aligned}$$

Thus, taking the norm on both sides and using its submultiplicativity yields

$$\begin{aligned}\left\|\mathbb{E}_{k-1}\left[\left(\boldsymbol{M}_k - \boldsymbol{M}_{k-1}\right)\left(\boldsymbol{M}_k - \boldsymbol{M}_{k-1}\right)^*\right]\right\| &\leq a^{2(n-1)} \cdot \mathbb{E}\|\boldsymbol{X}_k - \boldsymbol{X}\|^2 \\ &\leq a^{2(n-1)} s^2 \quad \text{almost surely.}\end{aligned}$$

Thus, we can define $v = n a^{2(n-1)} s^2$.

We can apply Lemma 2.3, and obtain for $t > 0$

$$\begin{aligned}\mathbb{P}\left(\left\|\boldsymbol{X}^{(n)} - \boldsymbol{X}^n\right\| \geq t\right) &\leq 2d \cdot \exp\left(\frac{-t^2/2}{v + Rt/3}\right) \\ &= 2d \cdot \exp\left(\frac{-t^2/2}{a^{n-1}(n a^{n-1} s^2 + rt/3)}\right).\end{aligned}$$

□

Some text about how $r \leq 2a$ holds by triangle formula, and $s \leq r \leq 2a$ similarly. Thus, in Theorem 2.1, one can infer a concentration result without conditions 2)i) and 2)ii).

Corollary 2.2 (Expectation bound for products)**.**
Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be independent random matrices of size $d \times d$. Assume that

1) there is a matrix $\boldsymbol{X}$ of size $d \times d$ such that for all $k = 1, \ldots, n$, we have $\mathbb{E}\left[\boldsymbol{X}_k\right] = \boldsymbol{X}$;
2) there are $a, r, s > 0$ such that for all $k = 1, \ldots, n$, we have
   i) $\|\boldsymbol{X}_k\| \leq a$ almost surely;
   ii) $\|\boldsymbol{X}_k - \boldsymbol{X}\| \leq r$ almost surely;
   iii) $\mathbb{E}\|\boldsymbol{X}_k - \boldsymbol{X}\| \leq s$.

Then,

$$\mathbb{E}\left\|\prod_{k=1}^{n} \boldsymbol{X}_k - \boldsymbol{X}^n\right\| \leq \sqrt{n} a^{(n-1)} s(\sqrt{2\ln(1+d)} + 2) + a^{n-1}\frac{r}{3}(2\ln(1+d) + 2).$$

Proof. Define $R = a^{n-1} r$ and $v = n a^{2(n-1)} s^2$. By the layer-cake representation theorem and Theorem 2.1, we have

$$\begin{aligned}\mathbb{E}\left\|\boldsymbol{X}^{(n)} - \boldsymbol{X}^n\right\| &= \int_0^\infty \mathbb{P}\left(\left\|\boldsymbol{X}^{(n)} - \boldsymbol{X}^n\right\| \geq t\right) \mathrm{d}t \\ &\leq \int_0^\infty \min\left\{1, 2d \cdot \exp\left(\frac{-t^2/2}{v + Rt/3}\right)\right\} \mathrm{d}t \\ &\leq \int_0^\mu 1 \,\mathrm{d}t + 2d \int_\mu^\infty \exp\left(\frac{-t^2/2}{v + Rt/3}\right) \mathrm{d}t\end{aligned}$$

We use the same argument as [Tro15] in the proof of Cor. 7.3.2 to further bound this expression. As their argument is instructive and short, we present their method on how to bound the integral of the exponential term. For $t \geq \sqrt{v}$, we have $\sqrt{v}(t/v) \geq 1$ and hence

$$\begin{aligned}\mathbb{E}\left\|\boldsymbol{X}^{(n)} - \boldsymbol{X}^n\right\| &\leq \mu + 2d\int_\mu^\infty e^{-t^2/(2v)}\,\mathrm{d}t + 2d\int_\mu^\infty e^{-3t/(2R}\,\mathrm{d}t \\ &\leq \mu + 2d\sqrt{v}\int_\mu^\infty (t/\sqrt{v})e^{-t^2/(2v)}\,\mathrm{d}t + 2d\int_\mu^\infty e^{-3t/(2R)}\,\mathrm{d}t \\ &= \mu + 2d\sqrt{v}e^{-\mu^2/(2v)} + 4/3de^{-3\mu/(2R)}\end{aligned}$$

Selecting $\mu = \sqrt{2v\ln(1+d)} + 2/3R\ln(1+d)$, we can find bounds for the terms in the above expression as

$$\begin{aligned}2d\sqrt{v}e^{-\mu^2/(2v)} &\leq 2d\sqrt{v}e^{-2v\ln(1+d)/(2v)} = 2d\sqrt{v}/(1+d) \leq 2\sqrt{v} \\ \frac{4}{3}dRe^{-3\mu/(2L)} &\leq \frac{4}{3}dRe^{-3(2/3)R\ln(1+d)/(2L)} = \frac{4}{3}dR/(1+d) \leq \frac{4}{3}R.\end{aligned}$$

Thus, we get

$$\begin{aligned}\mathbb{E}\left[\left\|\boldsymbol{X}^{(n)} - \boldsymbol{X}^n\right\|\right] &\leq \sqrt{2v\ln(1+d))} + \frac{2}{3}L\ln(1+d) + 2\sqrt{v} + \frac{4}{3}R \\ &= \sqrt{v}(\sqrt{2\ln(1+d)} + 2) + R/3(2\ln(1+d) + 2).\end{aligned}$$

□

Lemma 2.3 (Simplified Freedman's inequality for matrices, [CHKT21] Cor. 3.4, [Tro11a] Cor. 1.3)**.** Let $(\boldsymbol{M}_k)_{k=1,\dots,n}$ be a matrix martingale of size $d \times d$ adapted to some filtration $(\mathfrak{F}_k)_k$. Assume that

1) there is $R > 0$ such that for all $k = 1, \dots, n$ $\|\boldsymbol{M}_k - \boldsymbol{M}_{k-1}\| \leq R$ almost surely;

2) there is $v > 0$ such that $\|\sum_{k=1}^n \mathbb{E}_{k-1}\left[(\boldsymbol{M}_k - \boldsymbol{M}_{k-1})(\boldsymbol{M}_k - \boldsymbol{M}_{k-1})^*\right]\| \leq v$ almost surely.

Then, for $t > 0$

$$\mathbb{P}\left(\|\boldsymbol{M}_n - \boldsymbol{M}_0\| \geq t\right) \leq 2d \cdot \exp\left(\frac{-t^2/2}{v + Rt/3}\right).$$

Theorem 2.4 (Average-case concentration for products)**.**
Let $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ be independent random matrices of size $d \times d$. Assume that

1) there is a matrix $\boldsymbol{X}$ of size $d \times d$ such that for all $k = 1, \dots, n$, we have $\mathbb{E}\left[\boldsymbol{X}_k\right] = \boldsymbol{X}$;

2) there are $a, r, s > 0$ such that for all $k = 1, \dots, n$, we have

   i) $\|\boldsymbol{X}_k\| \leq a$ almost surely;

   ii) $\|\boldsymbol{X}_k - \boldsymbol{X}\| \leq r$ almost surely;

   iii) $\mathbb{E}\|\boldsymbol{X}_k - \boldsymbol{X}\| \leq s$.

Then, for any unit vector $\boldsymbol{u}$ of size d and $t > 0$

$$\mathbb{P}\left(\left\|\prod_{k=1}^n \boldsymbol{X}_k\boldsymbol{u} - \boldsymbol{X}^n\boldsymbol{u}\right\|_2 \geq t\right) \leq 2 \cdot \exp\left(\frac{-t^2/2}{a^{n-1}(na^{n-1}s^2 + rt/3)}\right).$$

Proof. This proof is conducted similarly to the proof of Theorem 2.1 using Lemma 2.6. □

In [Pin12], a powerful concentration result for vectors on smooth and separable Banach spaces is proved. It gives a probability bound for the supremum of the norm of martingales with expectation zero. Its utility lies in its generality as it allows us to find probability bounds which are similar to known concentration inequalities for the expression $\sup_{k\leq n}\|\boldsymbol{m}_k - \boldsymbol{m}_0\|$ instead of the smaller or at most equal expression $\|\boldsymbol{m}_n - \boldsymbol{m}_0\|$. The original result is more general as it considers vectors on a separable Banach space $\mathcal{X}$ which is $(2, D)$-smooth. This concept of smoothness is closely related to the uniform smoothness introduced earlier. A Banach space is called $(2, D)$-smooth if there is some $D > 0$ such that

$$\|\boldsymbol{x}+\boldsymbol{y}\|^2 + \|\boldsymbol{x}-\boldsymbol{y}\|^2 \leq 2\|\boldsymbol{x}\|^2 + 2D^2\|\boldsymbol{y}\|^2 \quad \text{for all } \boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}.$$

Since the parallelogram identity holds on Hilbert space, we have that $\mathcal{H}$ is $(2, 1)$-smooth. Therefore, we have the following version of Theorem 3.1 in [Pin12].

Lemma 2.5 (Vector concentration inequality on Hilbert spaces, [Pin12] Thm. 3.1)**.**
Let $\mathcal{H}$ be a separable Hilbert space. Let $(\boldsymbol{m}_k)_{k=1,\ldots,n}$ be a vector-valued martingale on $\mathcal{H}$ adapted to some filtration $(\mathfrak{F}_k)_k$. Set $\boldsymbol{d}_1 = \boldsymbol{m}_1$ and $\boldsymbol{d}_k = \boldsymbol{m}_k - \boldsymbol{m}_{k-1}$ for $k = 2\ldots, n$. Assume that $\mathbb{E}\left[e^{\lambda(\|\boldsymbol{d}_k\|)}\right] < \infty$ for all $k = 1, \ldots, n$. Then, for $t > 0$

$$\mathbb{P}\left(\sup_{k\leq n} \|\boldsymbol{m}_k - \boldsymbol{m}_0\| \geq t\right) \leq 2 \cdot \exp\left(-\lambda t + \sum_{k=1}^{n} \mathbb{E}_{k-1}\left[e^{\lambda \boldsymbol{d}_k} - \lambda \boldsymbol{d}_k - 1\right]\right).$$

In this work, we are concerned with random matrices that satisfy certain boundary conditions. In particular we impose conditions similar to those made in 2.1. This allows us to establish a Bernstein type inequality similar to Theorem 3.3 in [Pin12]. The following inequality is obtained using a similar approach as in the original proof of the Berstein inequality.

Lemma 2.6 (Vector Bernstein inequality on Hilbert spaces)**.**
Let $\mathcal{H}$ be a separable Hilbert space. Let $(\boldsymbol{m}_k)_{k=1,\ldots,n}$ be a vector-valued martingale on $\mathcal{H}$ adapted to some filtariton $(\mathfrak{F}_k)_k$. Assume that

1) there is $R > 0$ such that $\|\boldsymbol{m}_k - \boldsymbol{m}_{k-1}\| \leq R$ for all $k = 1, \ldots, n$;

2) there is $v > 0$ such that $\sum_{k=1}^{n} \mathbb{E}\left[\|\boldsymbol{m}_k - \boldsymbol{m}_{k-1}\|^2\right]$.

Then, for $t > 0$

$$\mathbb{P}\left(\sup_{k\leq n} \|\boldsymbol{m}_k - \boldsymbol{m}_0\| \geq t\right) \leq 2 \cdot \exp\left(-\frac{v}{R^2} h\left(\frac{tR}{v}\right)\right)$$

where $h(u) = (1+u)\ln(1+u) - u$. In particular, we have

$$\mathbb{P}\left(\sup_{k\leq n} \|\boldsymbol{m}_k - \boldsymbol{m}_0\| \geq t\right) \leq 2 \cdot \exp\left(\frac{-t^2/2}{v + Rt/3}\right).$$

One might expect we should be able to find concentration results in a similar way [HNWTW20] Theorem 2 is obtained via uniform smoothness. Our attempt to do so does not yield results for the concentration of random vectors that are not "close" to their expected value. Let us demonstrate why this is the case. Consider The setting of 2.4. Then, we would apply Markov's inequality to obtain

$$\mathbb{P}\left(\|\boldsymbol{X}^n\boldsymbol{u} - \boldsymbol{X}^n\boldsymbol{u}\|_2 \geq t\right) \leq \frac{\epsilon^q}{\mathbb{E}\left[\|\boldsymbol{X}^n\boldsymbol{u} - \boldsymbol{X}^{(n)}\boldsymbol{u}\|_2^q\right]} \tag{1}$$

for $q \geq 2$. It is natural to form the vector martingale $(\boldsymbol{m}_k)_k$ defined by

$$\boldsymbol{m}_k = \mathbb{E}_k\left[\boldsymbol{X}^{(n)}\boldsymbol{u}\right].$$

Now we are in a position to apply uniform smoothness similar to [CHKT21], [Che21], [HNWTW20]:

$$\begin{aligned}
\mathbb{E}\left[\left\|\boldsymbol{X}^n\boldsymbol{u}-\boldsymbol{X}^{(n)}\boldsymbol{u}\right\|_2^q\right]^{2/q} &= \mathbb{E}\left[\left\|\sum_{k=1}^{n}(\boldsymbol{m}_k-\boldsymbol{m}_{k-1})\right\|_2^q\right]^{2/q} \\
&= \mathbb{E}\left[\left\|\sum_{k=1}^{n-1}(\boldsymbol{m}_k-\boldsymbol{m}_{k-1})+(\boldsymbol{m}_n-\boldsymbol{m}_{n-1})\right\|_2^q\right]^{2/q} \\
&\le \mathbb{E}\left[\left\|\sum_{k=1}^{n-1}(\boldsymbol{m}_k-\boldsymbol{m}_{k-1})\right\|_2^q\right]^{2/q} + (q-1)\,\mathbb{E}\left[\|\boldsymbol{m}_n-\boldsymbol{m}_{n-1}\|_2^q\right]^{q/2} \\
&\le \dots \\
&\le (q-1)\sum_{k=1}^{n}\mathbb{E}\left[\|\boldsymbol{m}_k-\boldsymbol{m}_{k-1}\|_2^q\right]^{2/q}
\end{aligned}$$

It follows easily that

$$\mathbb{E}\left[\left\|\boldsymbol{m}_k-\boldsymbol{m}_{k-1}^q\right\|_2\right] \le \mathbb{E}\left[\left\|\boldsymbol{X}^{n-k}\right\|^q\|\boldsymbol{X}_k-\boldsymbol{X}\|^q\left\|\boldsymbol{X}^{(k-1)}\right\|^q\right] \le a^{q(n-1)}r^q.$$

We even see that the assumption $\mathbb{E}\left[\|\boldsymbol{X}_k-\boldsymbol{X}\|\right] \le s$ does not find application with this approach. Continuing the attempt, we find

$$\mathbb{E}\left[\left\|\boldsymbol{X}^n\boldsymbol{u}-\boldsymbol{X}^{(n)}\boldsymbol{u}\right\|_2^q\right]^{2/q} \le (q-1)na^{2(n-1)}r^2.$$

Plugging this into 1, we obtain

$$\mathbb{P}\left(\|\boldsymbol{X}^n\boldsymbol{u}-\boldsymbol{X}^n\boldsymbol{u}\|_2 \ge t\right) \le \left(\frac{nqa^{2(n-1)}r^2}{\epsilon^2}\right)^{q/2}.$$

We would choose

$$q = \frac{\epsilon^2}{ena^{2(n-1)}r^2}$$

to achieve a meaningful concentration result. However, q has to be larger than 2, otherwise the inequalities achieved with uniform smoothness are false. Notice that if we were to consider

$$\mathbb{P}\left(\frac{1}{n}\|\boldsymbol{X}^n\boldsymbol{u}-\boldsymbol{X}^n\boldsymbol{u}\|_2 \ge t\right)$$

we could restrict ourselves to n large enough, such that we can select

$$q = \frac{n\epsilon^2}{ea^{2(n-1)}r^2} \ge 2.$$

2.2 Perturbation of Identity

Theorem 2.7 (Perturbation of identity)**.**
Let $\boldsymbol{X}_1,\dots,\boldsymbol{X}_n$ be independent random matrices of size $d\times d$. Assume that

1) there is a matrix $\boldsymbol{X}$ of size $d\times d$ such that for all $k=1,\dots,n$, we have $\mathbb{E}\left[\boldsymbol{X}_k\right]$;
2) there are $a,r,s>0$ such that for all $k=1,\dots,n$, we have
   i) $\|\boldsymbol{X}_k\| \le a \le 1$ almost surely;
   ii) $\|\boldsymbol{X}_k-\boldsymbol{X}\| \le r$ almost surely;
   iii) $\mathbb{E}\|\boldsymbol{X}_k-\boldsymbol{X}\| \le s$.

Then, for $t > 0$,

$$\mathbb{P}\left(\left\|\prod_{k=1}^{n}\left(\mathbf{1}+\frac{1}{n}\boldsymbol{X}_k\right)-e^{\boldsymbol{X}}\right\|\geq t\right)\leq 2d\cdot\exp\left(\frac{-nt^2/8}{(e^a(e^a s^2+rt/3)}\right).$$

Proof. Use the notation $\boldsymbol{Y}_k = \left(\boldsymbol{I}+\frac{1}{n}\boldsymbol{X}_k\right)$, $\boldsymbol{Y}_j^{(k)} = \prod_{l=j}^{k}\boldsymbol{Y}_l$ and $\mathbb{E}_k[\,\cdot\,] = \mathbb{E}[\,\cdot \mid \boldsymbol{X}_1,\ldots,\boldsymbol{X}_k]$. Then, using the triangle formula

$$\left\|e^{\boldsymbol{X}}-\boldsymbol{Y}_1^{(n)}\right\|\leq\underbrace{\left\|e^{\boldsymbol{X}}-\mathbb{E}\left[\boldsymbol{Y}_1^{(n)}\right]\right\|}_{\text{deterministic bias}}+\underbrace{\left\|\mathbb{E}\left[\boldsymbol{Y}_1^{(n)}\right]-\boldsymbol{Y}_1^{(n)}\right\|}_{\text{fluctuation error}} \tag{2}$$

Notice that the deterministic part is non-random, and the fluctation error is random.

1) Deterministic bias: Use Lemma 2.9 and the indepedence of the random matrices to obtain

$$\begin{aligned}\left\|e^{\boldsymbol{X}}-\mathbb{E}\left[\boldsymbol{Y}_1^{(n)}\right]\right\| &= \left\|e^{\boldsymbol{X}}-\left(\boldsymbol{I}+\frac{1}{n}\boldsymbol{X}\right)^n\right\| \\ &\leq \frac{e^a}{n} \quad \text{almost surely.}\end{aligned}$$

. Thus for a given $t > 0$, if $n \geq 2e^a/n$, then $\left\|e^{\boldsymbol{X}}-\boldsymbol{Y}_1^{(n)}\right\| < t/2$ almost surely

2) Fluctuation error: We are looking to apply Theorem 2.1: Notice that $\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_n$ are independent random matrices and

   i) $\|\boldsymbol{Y}_k\| = \left\|1+\frac{1}{n}\boldsymbol{X}_k\right\| \leq 1+\frac{a}{n}$;
   ii) $\|\boldsymbol{Y}_k-\mathbb{E}\left[\boldsymbol{Y}_k\right]\| = \frac{1}{n}\|\boldsymbol{X}-\boldsymbol{X}\| \leq \frac{r}{n}$ almost surely;
   iii) $\mathbb{E}\|\boldsymbol{Y}_k-\mathbb{E}\left[\boldsymbol{Y}_k\right]\| = \frac{1}{n}\mathbb{E}\|\boldsymbol{X}_k-\boldsymbol{X}\| \leq \frac{s}{n}$ almost surely.

   Thus, we find for $t > 0$

$$\begin{aligned}\mathbb{P}\left(\left\|\mathbb{E}\left[\boldsymbol{Y}_1^{(n)}\right]-\boldsymbol{Y}_1^{(n)}\right\|\geq t\right) &\leq 2d\cdot\exp\left(\frac{-t^2/2}{(1+a/n)^{n-1}(n(1+a/n)^{n-1}(s/n)^2+(r/n)t/3)}\right) \\ &\leq 2d\cdot\exp\left(\frac{-nt^2/2}{(e^a(e^a s^2+rt/3)}\right)\end{aligned}$$

To bring both results together use the union bound and find that for $n \geq 2e^{(a)}/n$

$$\begin{aligned}\mathbb{P}\left(\left\|e^{\boldsymbol{X}}-\boldsymbol{Y}_1^{(n)}\right\|\geq t\right) &\leq \mathbb{P}\left(\left\|e^{\boldsymbol{X}}-\mathbb{E}\left[\boldsymbol{Y}_1^{(n)}\right]\right\|\geq t/2\right)+\mathbb{P}\left(\left\|\boldsymbol{Y}_1^{(n)}-\mathbb{E}\left[\boldsymbol{Y}_1^{(n)}\right]\right\|\geq t/2\right) \\ &= 0+\mathbb{P}\left(\left\|\boldsymbol{Y}_1^{(n)}-\mathbb{E}\left[\boldsymbol{Y}_1^{(n)}\right]\right\|\geq t/2\right) \\ &\leq 2d\cdot\exp\left(\frac{-nt^2/8}{(e^a(e^a s^2+rt/3)}\right)\end{aligned}$$

□

We can use the theorem about concentration of products to find a bound for the random fluctuation and simplify the bound afterwards.

Corollary 2.8 (Perturbation of identity on infinite dimensions)**.**
Let $\mathcal{H}$ be a separable Hilbert space. Let $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n$ be independent random trace-class operators on $\mathcal{H}$. Assume that

1) there is a trace-class operator $\boldsymbol{X}$ such that for all $k = 1,\ldots,n$, we have $\mathbb{E}\left[\boldsymbol{X}_k\right] = \boldsymbol{X}$;

2) there is a trace-class operator $\boldsymbol{V}$ such that for all $k = 1,\ldots,n$, we have $\boldsymbol{V} \succeq \mathbb{E}\left[\boldsymbol{X}^*\boldsymbol{X}\right]$;

3) there are $a, r, s > 0$ such that for all $k = 1,\ldots,n$, we have

i) $\|\boldsymbol{X}_k\| \le a \le 1$ almost surely;

ii) $\|\boldsymbol{X}_k - \boldsymbol{X}\| \le r$ almost surely;

iii) $\mathbb{E}\|\boldsymbol{X}_k - \boldsymbol{X}\| \le s$.

Then, for $t > 0$,

$$\mathbb{P}\left(\left\|\prod_{k=1}^{n}\left(\mathbf{1} + \frac{1}{n}\boldsymbol{X}_k\right) - e^{\boldsymbol{X}}\right\| \ge t\right) \le 4\,\mathrm{int}\dim(\boldsymbol{V}) \cdot \exp\left(\frac{-nt^2/8}{(e^a(e^a s^2 + rt/3)}\right).$$

Proof. The proof is identical to the proof of Theorem 2.7 using Theorem (!!! intrinsic Bernstein inequality) □

Lemma 2.9.
Let $\boldsymbol{T} \in \mathcal{L}(\mathcal{H})$ be a contraction, i.e. $\|\boldsymbol{T}\| \le 1$. (De we need that $\boldsymbol{T}$ is self-adjoint?) Then,

$$\left\|\left(I + \frac{1}{n}\boldsymbol{T}\right)^n - e^{\boldsymbol{T}}\right\| \le \frac{e^{\|\boldsymbol{T}\|}}{n}.$$

Proof. Let $t \in [-1, 1]$. Then, $\left(1 + \frac{t}{n}\right)^n < e^t < \left(1 + \frac{t}{n}\right)^{n+1}$. This follows considering $f(\tau) = \left(1 + \frac{\tau}{n}\right)^n$, $f(\tau) = e^\tau$, $f_2(\tau) = \left(1 + \frac{\tau}{n}\right)^{n+1}$ and noting that

$$\begin{aligned} f_1(0) &= 1, \quad f_1'(\tau) < f_1(\tau) \\ f(0) &= 1, \quad f'(\tau) = f(\tau) \\ f_2(0) &= 1, \quad f_2'(\tau) > f_2(\tau) \end{aligned}$$

Therefore,

$$\left|e^t - \left(1 + \frac{t}{n}\right)^n\right| \le \left(1 + \frac{t}{n}\right)^{n+1} - \left(1 + \frac{t}{n}\right)^n = \left(1 + \frac{t}{n}\right)^n \left|\frac{t}{n}\right| \le \frac{e^{\|\boldsymbol{T}\|}}{n}.$$

Now apply continuous functional calculus to obtain the result. □

Corollary 2.10 (Expectation bound for perturbation of identity)**.**
Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be independent random matrices of size $d \times d$. Assume that

1) there is a matrix $\boldsymbol{X}$ of size $d \times d$ such that for all $k = 1, \ldots, n$, we have $\mathbb{E}\left[\boldsymbol{X}_k\right]$;

2) there are $a, r, s > 0$ such that for all $k = 1, \ldots, n$, we have

   i) $\|\boldsymbol{X}_k\| \le a \le 1$ almost surely;

   ii) $\|\boldsymbol{X}_k - \boldsymbol{X}\| \le r$ almost surely;

   iii) $\mathbb{E}\|\boldsymbol{X}_k - \boldsymbol{X}\| \le s$.

Then,
$$\left\|e^{\boldsymbol{X}} - \boldsymbol{Y}_1^{(n)}\right\| \le \frac{1}{\sqrt{n}}\left[e^a s(\sqrt{2\ln(1+d)} + 2)\right] + \frac{e^a}{n}\left[\frac{r}{3}(2\ln(1+d) + 2) + 1\right].$$

Proof. Notice we can deduce the inequality as in 2 to obtain

$$\left\|e^{\boldsymbol{X}} - \boldsymbol{Y}_1^{(n)}\right\| \le \left\|e^{\boldsymbol{X}} - \mathbb{E}\left[\boldsymbol{Y}_1^{(n)}\right]\right\| + \left\|\mathbb{E}\left[\boldsymbol{Y}_1^{(n)}\right] - \boldsymbol{Y}_1^{(n)}\right\|.$$

Then, the first part can be bounded by e^a/n. Notice that $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$ are independent random matrices and

i) $\|\boldsymbol{Y}_k\| = \left\|1 + \frac{1}{n}\boldsymbol{X}_k\right\| \le 1 + \frac{a}{n}$;

ii) $\|\boldsymbol{Y}_k - \mathbb{E}\left[\boldsymbol{Y}_k\right]\| = \frac{1}{n}\|\boldsymbol{X} - \boldsymbol{X}\| \le \frac{r}{n}$ almost surely;

iii) $\mathbb{E}\|\boldsymbol{Y}_k - \mathbb{E}\left[\boldsymbol{Y}_k\right]\| = \frac{1}{n}\mathbb{E}\|\boldsymbol{X}_k - \boldsymbol{X}\| \le \frac{s}{n}$ almost surely.

Thus, we can apply Corollary 2.2 to find

$$\begin{aligned}\mathbb{E}\left\|\boldsymbol{Y}_1^{(n)} - \boldsymbol{Y}^n\right\| &\leq \sqrt{n}(1+a/n)^{(n-1)}(s/n)(\sqrt{2\ln(1+d)}+2) + (1+a/n)^{n-1}(r/n)/3(2\ln(1+d)+2) \\ &= 1/\sqrt{n}e^a s(\sqrt{2\ln(1+d)}+2) + e^a(r/n)/3(2\ln(1+d)+2)\end{aligned}$$

In total, we have

$$\left\|e^{\boldsymbol{X}} - \boldsymbol{Y}_1^{(n)}\right\| \leq \frac{1}{\sqrt{n}}\left[e^a s(\sqrt{2\ln(1+d)}+2)\right] + \frac{e^a}{n}\left[\frac{r}{3}(2\ln(1+d)+2)+1\right].$$

□

A similar expectation bound holds for trace-class operators on separable Hilbert spaces. We only have to account for the change in some details due to changing d to $2\,\mathrm{int}\dim(\boldsymbol{V})$.

2.2.1 Average-Case Concentration for Perturbation of Identity

Theorem 2.11 (Perturbation of identity)**.**
Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be independent random matrices of size $d \times d$. Assume that

1) there is a matrix $\boldsymbol{X}$ of size $d \times d$ such that for all $k = 1, \ldots, n$, we have $\mathbb{E}[\boldsymbol{X}_k] = \boldsymbol{X}$;
2) there are $a, r, s > 0$ such that for all $k = 1, \ldots, n$, we have
   i) $\|\boldsymbol{X}_k\| \leq a$ almost surely;
   ii) $\|\boldsymbol{X}_k - \boldsymbol{X}\| \leq r$ almost surely;
   iii) $\mathbb{E}\|\boldsymbol{X}_k - \boldsymbol{X}\| \leq s$.

Then, for any unit vector u of size d and $t > 0$,

$$\mathbb{P}\left(\left\|\prod_{k=1}^{n}\left(\mathbf{1} + \frac{1}{n}\boldsymbol{X}_k\right)u - e^{\boldsymbol{X}}u\right\|_2 \geq t\right) \leq 2d \cdot \exp\left(\frac{-nt^2/8}{(e^a(e^a s^2 + rt/3)}\right).$$

2.3 Concentration for Products of Unitaries

References

[Che21] Chi-Fang Chen. Concentration of otoc and lieb-robinson velocity in random hamiltonians, 2021.

[CHKT21] Chi-Fang Chen, Hsin-Yuan Huang, Richard Kueng, and Joel A. Tropp. Concentration for random product formulas. *PRX Quantum*, 2(4), October 2021.

[Gol65] Sidney Golden. Lower bounds for the helmholtz function. *Phys. Rev.*, 137:B1127–B1128, Feb 1965.

[HNWTW20] De Huang, Jonathan Niles-Weed, Joel A. Tropp, and Rachel Ward. Matrix concentration for products, 2020.

[Lie73] Elliott H Lieb. Convex trace functions and the wigner-yanase-dyson conjecture. *Advances in Mathematics*, 11(3):267–288, 1973.

[Pin12] Iosif Pinelis. Optimum bounds for the distributions of martingales in banach spaces, 2012.

[Tho65] C. J. Thompson. Inequality with applications in statistical mechanics. *Journal of Mathematical Physics*, 6:1812–1813, 1965.

[Tro11a] Joel A. Tropp. Freedman's inequality for matrix martingales, 2011.

[Tro11b] Joel A. Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics*, 12(4):389–434, August 2011.

[Tro15] Joel A. Tropp. An introduction to matrix concentration inequalities, 2015.