1 Golden-Thompson

Recall the Golden-Thompson inequality, employed in the previous lecture:

Lemma 1.1. *For all symmetric matrices $A, B \in \mathbb{M}_d$, it holds that*

$$\mathrm{Tr}(e^{A+B}) \leqslant \mathrm{Tr}(e^A e^B).$$

Recall that $e^X = \sum_{n \geqslant 0} \frac{X^n}{n!}$, and note that

$$(A+B)^n = \sum_{i_1, i_2, \ldots, i_n \in \{0,1\}} A^{i_1} B^{1-i_1} A^{i_2} B^{1-i_2} \cdots A^{i_n} B^{1-i_n}$$

is a uniform sum over all degree-n interleavings of A and B. To obtain the degree-n terms in $e^A e^B$, one takes every product occuring in this sum and sorts it so that all the copies of A come first: $e^A e^B$ only contains products of the form $A^j B^{n-j}$ for $0 \leqslant j \leqslant n$.

Thus, intuitively, $\mathrm{Tr}(e^{A+B}) \leqslant \mathrm{Tr}(e^A e^B)$ asserts that the largest correlations occur when the A and B terms are grouped together. How might we prove this?

1.1 The Frobenius inner product

A hint comes from the Cauchy-Schwarz inequality. Define the *Frobenius inner product* of two matrices $A, B \in \mathbb{M}_d$ by

$$(A, B) \mapsto \mathrm{Tr}(A^T B) = \sum_{ij} A_{ij} B_{ij}.$$

(Similarly for $A, B \in \mathbb{M}_d(\mathbb{C})$, one uses $\mathrm{Tr}(A^* B)$.) As the final expression shows, this is just the standard inner product on the "vectorizations" of A and B (i.e., when considering them as d^2-dimensional vectors).

Let us correspondingly define the Frobenius norm (aka the Hilbert-Schmidt norm, aka the Schatten 2-norm) of a matrix:

$$\|A\|_2 := \mathrm{Tr}\left(A^T A\right)^{1/2}.$$

Recall that every matrix $A \in \mathbb{M}_d$ has a singular-value decomposition $A = \sum_{i=1}^d \sigma_i u_i v_i^T$, where $\sigma_1, \ldots, \sigma_d \geqslant 0$, and each of $\{u_i\}, \{v_i\}$ forms an orthonormal basis of $\mathbb{R}^d$. In this case,

$$A^T A = \sum_i \sigma_i^2 v_i v_i^T,$$

hence we also have $\|A\|_2 = \|(\sigma_1, \ldots, \sigma_d)\|_2$.

1.2 Sorting products

Assume now that $A, B \in \mathbb{M}_d$ are symmetric, and then applying the Cauchy-Schwarz inequality gives

$$\mathrm{Tr}((AB)^2) \leqslant \|AB\|_2^2 = \mathrm{Tr}((AB)^T AB) = \mathrm{Tr}(B^T A^T AB) = \mathrm{Tr}(BA^2 B) = \mathrm{Tr}(A^2 B^2), \tag{1.1}$$

where the last equality uses cyclicity of the trace. Let's try the fourth power:

$$\begin{aligned}\mathrm{Tr}((AB)^4) = \mathrm{Tr}((AB)^2(AB)^2) \leqslant \|(AB)^2\|_2^2 &= \mathrm{Tr}(BABA\ ABAB) \\ &= \mathrm{Tr}([AB(AB)^T][(AB)^T AB]) \\ &\leqslant \|AB(AB)^T\|_2 \|(AB)^T AB\|_2 \\ &= \|AB(AB)^T\|_2^2 \\ &= \mathrm{Tr}((AB^2A)^2) = \mathrm{Tr}((A^2B^2)^2),\end{aligned}$$

where the last equality uses cyclicity of the trace. Now we can apply (1.1) (with the substitution $A \to A^2, B \to B^2$), yielding

$$\mathrm{Tr}((AB^4) \leqslant \mathrm{Tr}((A^2B^2)^2) \leqslant \mathrm{Tr}(A^4B^4).$$

This gives one faith that such a relation holds more generally; we will prove the following.

Lemma 1.2 (Distentangling Lemma)**.** *For every integer $k \geqslant 1$ and $U, V \in \mathbb{M}_d(\mathbb{C})$ Hermitian, it holds that*

$$\mathrm{Tr}((UV)^{2^k}) \leqslant \mathrm{Tr}(U^{2^k} V^{2^k}).$$

1.2.1 The Lie-Trotter product formula

Now let us see how to employ the sorting lemma (Lemma 1.2) to prove the Golden-Thompson inequality. If we take $U := e^{A/p}$ and $V := e^{B/p}$ for some $p = 2^k$, then Lemma 1.2 gives

$$\mathrm{Tr}\left((e^{A/p}e^{B/p})^p\right) \leqslant \mathrm{Tr}(e^A e^B). \tag{1.2}$$

Now we can employ the Lie-Trotter formula which asserts that for any matrices $A, B \in \mathbb{M}_d$,

$$e^{A+B} = \lim_{N\to\infty} \left(e^{A/N} e^{B/N}\right)^N. \tag{1.3}$$

Thus taking $p \to \infty$ in (1.2) yields Lemma 1.1.

Proof of (1.3). Denote $U := e^{(A+B)/N}$ and $V := e^{A/N}e^{B/N}$. Then using $e^X = \sum_{n\geqslant 0} X^n/n!$, we have

$$\begin{aligned} U &= I + \frac{A+B}{N} + \frac{(A+B)^2}{2N^2} + \cdots \\ V &= I + \frac{A+B}{N} + \frac{A^2+B^2+2AB}{2N^2} + \cdots, \end{aligned}$$

so U and V agree up to first order, hence

$$\|U - V\| \leqslant O(1/N^2). \tag{1.4}$$

where the $O(\cdot)$ notation hides a constant possibly depending on A and B (but not on N). Note also that $\|U\|, \|V\| \leqslant e^{(\|A\|+\|B\|)/N}$.

Using both these facts and the triangle inequality give

$$\|U^N - V^N\| = \sum_{k=0}^{N-1} \left\|U^{k+1}V^{N-k-1} - U^k V^{N-k}\right\|$$

$$
\begin{aligned}
&= \sum_{k=0}^{N-1} \left\| U^k (U - V) V^{N-k-1} \right\| \\
&\leqslant \|U - V\| \sum_{k=0}^{N-1} \|U^k\| \cdot \|V^{N-k-1}\| \\
&\leqslant N \|U - V\| e^{\|A\|+\|B\|} \\
&\overset{(1.4)}{\leqslant} O(1/N),
\end{aligned}
$$

where we used the fact that $\|ST\| \leqslant \|S\| \cdot \|T\|$ holds for all $S, T \in \mathbb{M}_d$. As $U^N = e^{A+B}$ and $V^N = (e^{A/N} e^{B/N})^N$, this completes the proof. □

1.2.2 Disentangling

Our proof of Lemma 1.2 will follow an argument of Dyson (1964). Note that Cauchy-Schwarz gives $\mathrm{Tr}(A^2) \leqslant \mathrm{Tr}(A^* A)$ for all $A \in \mathbb{M}_d(\mathbb{C})$. Let us prove the following generalization.

Lemma 1.3. *Consider $A \in \mathbb{M}_d(\mathbb{C})$, and suppose that $A_i \in \{A, A^*\}$ for each $i = 1, 2, \ldots, 2n$. Then,*

$$
|\mathrm{Tr}(A_1 A_2 \cdots A_{2n})| \leqslant \mathrm{Tr}\left((A^* A)^n\right).
$$

Proof. We may clearly assume that $A \neq A^*$. Let $\mathcal{P}_n$ denote the space of such products $P = A_1 A_2 \cdots A_{2n}$. Define the number of *transitions* in P as $\#\left\{ i \in \{1, 2, \ldots, 2n\} : A_{i \bmod 2n} \neq A_{(i+1) \bmod 2n} \right\}$, i.e., the number of times in the cyclic order we see AA^* or $A^* A$ occur in P.

Let $P = A_1 A_2 \cdots A_{2n}$ denote a maximizer of $|\mathrm{Tr}(P)|$ among $P \in \mathcal{P}_n$. If the number of transitions in P is $2n$, then $P = (A^* A)^n$ or $P = (AA^*)^n$, and we are done. Otherwise, there is some adjacent pair of symbols that are equal; by a cyclic permutation, we may assume that $A_n = A_{n+1}$.

Denote $Q := A_1 \cdots A_n$ and $R := A_{n+1} \cdots A_{2n}$, as well as $P' = Q^* Q$ and $P'' = R^* R$ so that $P', P'' \in \mathcal{P}_n$. By Cauchy-Schwarz, we have

$$
|\mathrm{Tr}(P)|^2 \leqslant |\mathrm{Tr}(Q^* Q)| \cdot |\mathrm{Tr}(R^* R)| = |\mathrm{Tr}(P')| \cdot |\mathrm{Tr}(P'')|.
$$

By maximality of $|\mathrm{Tr}(P)|$, we have $|\mathrm{Tr}(P)| = |\mathrm{Tr}(P')| = |\mathrm{Tr}(P'')|$. We will argue that one of P' or P'' has more transitions than P, and therefore by induction there exists a maximizer of $|\mathrm{Tr}(P)|$ with $2n$ transitions, completing the proof.

Indeed, the transitions in P are made up of three types: Those that occur within Q, those that occur within R, and possibly one transition from A_{2n} to A_1, hence

$$
N_P \leqslant n_Q + n_R + 1.
$$

Moreover, we have $N_{P'} \geqslant 2n_Q + 2$ and $N_{P''} \geqslant 2n_R + 2$. Let us prove the first inequality, since the second is identical. Every transition within Q induces two transitions in P', one in Q^* and one in Q. There are also two new transitions: One from the end of Q^* to the beginning of Q, and one from the end of Q to the beginning of Q^*. We conclude that $(N_{P'} + N_{P''})/2 > N_P$, hence one of P' or P'' has more transitions than P. □

We can now prove Lemma 1.2.

Proof of Lemma 1.2. Recall that U, V are Hermitian. Define $A = UV$ so that Lemma 1.3 gives

$$\mathrm{Tr}((UV)^{2k}) \leqslant \mathrm{Tr}((V^*U^*UV)^{2^{k-1}}) = \mathrm{Tr}((VU^2V)^{2^{k-1}}) = \mathrm{Tr}((U^2V^2)^{2^{k-1}}),$$

where the last equality uses cyclicity of the trace. Continuing inductively gives

$$\mathrm{Tr}((U^2V^2)^{2^{k-1}}) \leqslant \mathrm{Tr}((U^4V^4)^{2^{k-2}}) \leqslant \cdots \leqslant \mathrm{Tr}(U^{2^k}V^{2^k}) \qquad \square.$$

1.3 A Hölder product formula

Let us prove a generalization of Lemma 1.3. For this, we define the *Schatten p-norm* of a matrix $A \in \mathbb{M}_d(\mathbb{C})$: For any $p \geqslant 1$, define

$$\|A\|_p := (\mathrm{Tr}(|A|^p))^{1/p} = \left(\mathrm{Tr}((A^*A)^{p/2})\right)^{1/p}.$$

The operator norm $\|A\| = \|A\|_\infty$ is the limiting case as $p \to \infty$. One can see that, as for the 2-norm, if $\sigma_1, \ldots, \sigma_d \geqslant 0$ are the singular values of A, then

$$\|A\|_p = \|(\sigma_1, \ldots, \sigma_d)\|_p.$$

Lemma 1.4. *For any integer $k \geqslant 1$ and $A_1, \ldots, A_{2^k} \in \mathbb{M}_d(\mathbb{C})$, it holds that*

$$|\mathrm{Tr}\,(A_1A_2\cdots A_{2^k})| \leqslant \|A_1\|_{2^k}\|A_2\|_{2^k}\cdots\|A_{2^k}\|_{2^k}.$$

Proof. The proof is by induction on k. The case $k = 1$ is Cauchy-Schwarz.

Consider now $k > 1$. The inductive hypothesis yields

$$|\mathrm{Tr}\,(A_1A_2\cdots A_{2^k})| \leqslant \|A_1A_2\|_{2^{k-1}}\|A_3A_4\|_{2^{k-1}}\cdots\|A_{2^k-1}A_{2^k}\|_{2^{k-1}}. \tag{1.5}$$

Now use the definition of the Schatten 2^{k-1}-norm to write

$$\|A_1A_2\|_{2^{k-1}}^{2^{k-1}} = \mathrm{Tr}\left(((A_1A_2)^*(A_1A_2))^{2^{k-2}}\right) = \mathrm{Tr}\left((A_2^*A_1^*A_1A_2)^{2^{k-2}}\right) = \mathrm{Tr}\left(A_1^*A_1A_2^*A_2A_1^*A_1\cdots A_2^*A_2\right),$$

where in the last equality we have used the cyclic property of the trace to move one copy of A_2^* from the head to the tail of the product. Applying the inductive hypothesis again yields

$$\mathrm{Tr}\left(A_1A_1^*A_2A_2^*\cdots A_2A_2^*\right) \leqslant \prod_{j=1}^{2^{k-1}} \|A_1A_1^*\|_{2^{k-1}}\|A_2A_2^*\|_{2^{k-1}} = \|A_1\|_{2^{k-1}}^{2^k}\|A_2\|_{2^{k-1}}^{2^k},$$

where we have observed that $\|A_1A_1^*\|_{2^{k-1}}^{2^{k-1}} = \mathrm{Tr}\left((A_1A_1^*A_1A_1^*)^{2^{k-2}}\right) = \mathrm{Tr}((A_1A_1^*)^{2^{k-1}}) = \|A_1\|_{2^k}^{2^k}$, and similarly for A_2. Therefore we have

$$\|A_1A_2\|_{2^{k-1}} \leqslant \|A_1\|_{2^k}\|A_2\|_{2^k}.$$

Since this holds also for every pair A_iA_{i+1}, using it in (1.5) yields

$$|\mathrm{Tr}\,(A_1A_2\cdots A_{2^k})| \leqslant \|A_1\|_{2^k}\|A_2\|_{2^k}\cdots\|A_{2^k}\|_{2^k},$$

as desired. $\square$

In analogy with the scalar case, we might look to prove a generalization of Lemma 1.4: For any $p_1, p_2, \ldots, p_n > 0$ such that $\frac{1}{p_1} + \cdots + \frac{1}{p_n} = 1$,

$$|\mathrm{Tr}(A_1 A_2 \cdots A_n)| \leqslant \|A_1\|_{p_1} \|A_2\|_{p_2} \cdots \|A_n\|_{p_n}.$$

To prove this, fix some $m \geqslant 1$ and define $N := 2^m$, $k_i := \lfloor N/p_i \rfloor$ for each $i = 1, \ldots, n$. Then we have

$$\begin{aligned}
|\mathrm{Tr}(A_1 A_2 \cdots A_n)| &= \left| \mathrm{Tr}\left(\prod_{j=1}^{k_1} A_1^{1/k_1} \cdot \prod_{j=1}^{k_2} A_2^{1/k_2} \cdots \prod_{j=1}^{k_n} A_n^{1/k_n} I^{N-(k_1+\cdots+k_n)} \right) \right| \\
&\leqslant \|A_1^{1/k_1}\|_N^{k_1} \cdots \|A_n^{1/k_n}\|_N^{k_n} \cdot \|I\|_N^{N-(k_1+\cdots+k_n)}.
\end{aligned}$$

Note that

$$\|A^{1/k}\|_N^k = \mathrm{Tr}(|A|^{N/k})^{k/N} = \|A\|_{N/k},$$

and

$$\|I\|_N^{N-(k_1+\cdots+k_n)} = d^{1-(k_1+\cdots+k_n)/N},$$

hence

$$|\mathrm{Tr}(A_1 A_2 \cdots A_n)| \leqslant \|A_1\|_{2^m/k_1} \cdots \|A_n\|_{2^m/k_n} \cdot d^{1-(k_1+\cdots+k_n)/2^m}.$$

As $m \to \infty$, we have $2^m/k_i \to p_i$ and $(k_1 + \cdots + k_n)/2^m \to 1$, completing the proof.

1.4 Discussion

Perhaps that all seemed a bit mysterious. While "non-interleaved correlations are the largest" makes intuitive sense, why does something clean like Lemma 1.1 hold? Say that a norm $|||\cdot|||$ on $\mathbb{M}_d(\mathbb{C})$ is *unitarily invariant* if $|||UAV||| = |||A|||$ for all $A \in \mathbb{M}_d(\mathbb{C})$ and U, V unitary. (We will study unitarily invariant norms more in the next lecture.)

The trace norm $A \mapsto \mathrm{Tr}((A^*A)^{1/2})$ is such a norm (as are all Schatten p-norms for $p \in [1, \infty]$). An analog of Lemma 1.1 holds for every unitarily invariant norm: If $A, B \in \mathbb{M}_d(\mathbb{C})$, then

$$|||e^{A+B}||| \leqslant |||e^{A/2} e^B e^{A/2}|||. \tag{1.6}$$

Weak majorization. Inequality (1.6) holding for every unitarily invariant norm is equivalent to the statement that

$$e^{A+B} \prec_w e^{A/2} e^B e^{A/2},$$

where for two matrices $X, Y \in \mathbb{M}_d(\mathbb{C})$ with singular values $\sigma_1(X) \geqslant \cdots \geqslant \sigma_d(X)$ and $\sigma_1(Y) \geqslant \cdots \geqslant \sigma_d(Y)$, the notation $X \prec_w Y$ means that

$$\sigma_1(X) + \cdots + \sigma_k(X) \leqslant \sigma_1(Y) + \cdots + \sigma_k(Y), \quad \forall 1 \leqslant k \leqslant d.$$

In general, this inequality is related to similar sorts of "non-interleaved correlations are the largest" inequalities. For instance, it holds that for every pair of PSD matrices $A, B \in \mathbb{M}_d(\mathbb{C})$ and any Hermitian $X \in \mathbb{M}_d(\mathbb{C})$:

$$|||A^{1/2} X B^{1/2}||| \leqslant \left|\left|\left| \int_0^1 A^t X B^{1-t}\, dt \right|\right|\right| \leqslant \left|\left|\left| \frac{AX + XB}{2} \right|\right|\right|.$$

This is one possible analog of the classical AM-LM-GM inequality: For all $a, b \geqslant 0$, it holds that

$$\sqrt{ab} \leqslant \int_0^1 a^t b^{1-t}\, dt \leqslant \frac{a+b}{2}.$$

(The less familiar quantity in the middle is the "logarithmic mean" and equals $\frac{a-b}{\log a - \log b}$.) We will discuss such concepts further when we study matrix means.

Here is another example:

Theorem 1.5 (Lieb-Thirring trace inequality)**.** *For all $A, B \succeq 0$ and $t \geqslant 1$, it holds that*

$$\operatorname{Tr}\left[(B^{1/2}AB^{1/2})^t\right] \leqslant \operatorname{Tr}\left[A^t B^t\right]$$