

21.12 Golden–Thompson

For numbers, $e^{a+b} = e^a e^b$. For matrices this fails the moment $\mathbf{A}$ and $\mathbf{B}$ stop commuting, and Chapter 10 §09 was careful to say so: Theorem 10.9.4 (d) buys the identity only at the price of $\mathbf{AB} = \mathbf{BA}$. The question this section answers is what survives without that price. The answer is that the **trace** still knows the right inequality, in one direction, and that the direction is the useful one.

The section also closes the chapter, and Part V with it. After Golden–Thompson come two of its relatives — Klein’s inequality and the nonnegativity of the relative entropy — a counterexample showing that three matrices are one too many, and one theorem that is quoted rather than proved.

21.12.1 Sums in the exponent, products in the answer

A product $e^{\mathbf{A}}e^{\mathbf{B}}$ is not $e^{\mathbf{A}+\mathbf{B}}$, but it is not far off either: to first order in $\mathbf{A}$ and $\mathbf{B}$ the two agree, since both are $\mathbf{I} + \mathbf{A} + \mathbf{B} + \dots$, and the discrepancy is quadratic. So if the exponents are made small by a factor $1/k$ and the resulting error is repaired k times over, the discrepancy should vanish. It does, and this is the one tool that turns a sum in the exponent into a product.

Lemma 21.12.1. Lie’s Product Formula

Let $\mathbf{A}, \mathbf{B} \in M_n(\mathbb{C})$. Then

$$e^{\mathbf{A}+\mathbf{B}} = \lim_{k \rightarrow \infty} (e^{\mathbf{A}/k} e^{\mathbf{B}/k})^k,$$

the limit taken entrywise, and also in any norm on $M_n(\mathbb{C})$.

Idea. Write $\mathbf{S}_k = e^{(\mathbf{A}+\mathbf{B})/k}$ and $\mathbf{T}_k = e^{\mathbf{A}/k} e^{\mathbf{B}/k}$. Then $\mathbf{S}_k^k$ is exactly $e^{\mathbf{A}+\mathbf{B}}$, so the whole task is to compare $\mathbf{S}_k^k$ with $\mathbf{T}_k^k$. Two estimates do it. Both $\mathbf{S}_k$ and $\mathbf{T}_k$ have norm at most $e^{(\|\mathbf{A}\|+\|\mathbf{B}\|)/k}$, so no power of either can be large; and $\mathbf{S}_k - \mathbf{T}_k$ is of size $1/k^2$, because the two agree through the linear term. A telescoping identity turns a difference of k -th powers into k copies of $\mathbf{S}_k - \mathbf{T}_k$, and $k \cdot 1/k^2$ tends to 0.

Proof. Fix the spectral norm $\|\cdot\|_2$, which is submultiplicative by Theorem 16.3.2 (d). Put $a = \|\mathbf{A}\|_2$, $b = \|\mathbf{B}\|_2$, and $c = a + b$, and let k be an integer with $k \geq \max(1, c)$. Write

$$\mathbf{S}_k = e^{(\mathbf{A}+\mathbf{B})/k}, \quad \mathbf{T}_k = e^{\mathbf{A}/k} e^{\mathbf{B}/k}.$$

Step 1. $\mathbf{S}_k^k = e^{\mathbf{A}+\mathbf{B}}$. The matrix $(\mathbf{A} + \mathbf{B})/k$ commutes with itself, so Theorem 10.9.4 (d) gives $e^{j(\mathbf{A}+\mathbf{B})/k} e^{(\mathbf{A}+\mathbf{B})/k} = e^{(j+1)(\mathbf{A}+\mathbf{B})/k}$ for every $j \geq 0$. Induction on j yields $\mathbf{S}_k^j = e^{j(\mathbf{A}+\mathbf{B})/k}$, and $j = k$ is the claim.

Step 2. Norm bounds. By Theorem 16.6.1 (a), $\|\mathbf{S}_k\|_2 \leq e^{\|\mathbf{A}+\mathbf{B}\|_2/k} \leq e^{c/k}$, using $\|\mathbf{A} + \mathbf{B}\|_2 \leq c$; and $\|\mathbf{T}_k\|_2 \leq e^{a/k} e^{b/k} = e^{c/k}$ by submultiplicativity. Since $k \geq c$, both are at most e .

Step 3. $\mathbf{S}_k - \mathbf{T}_k$ is of size $1/k^2$. For any $\mathbf{X} \in M_n(\mathbb{C})$ put $\mathbf{R}(\mathbf{X}) = e^{\mathbf{X}} - \mathbf{I} - \mathbf{X}$, so that

Theorem 16.6.1 (c) gives

$$\|\mathbf{R}(\mathbf{X})\|_2 \leq \frac{1}{2} \|\mathbf{X}\|_2^2 e^{\|\mathbf{X}\|_2}.$$

Multiplying out $\mathbf{T}_k = (\mathbf{I} + \frac{\mathbf{A}}{k} + \mathbf{R}(\frac{\mathbf{A}}{k}))(\mathbf{I} + \frac{\mathbf{B}}{k} + \mathbf{R}(\frac{\mathbf{B}}{k}))$, and using $e^{\mathbf{B}/k} = \mathbf{I} + \frac{\mathbf{B}}{k} + \mathbf{R}(\frac{\mathbf{B}}{k})$ to collect the three terms that begin with $\mathbf{R}(\frac{\mathbf{A}}{k})$,

$$\mathbf{T}_k = \mathbf{I} + \frac{\mathbf{A} + \mathbf{B}}{k} + \mathbf{R}\left(\frac{\mathbf{B}}{k}\right) + \frac{\mathbf{AB}}{k^2} + \frac{\mathbf{A}}{k} \mathbf{R}\left(\frac{\mathbf{B}}{k}\right) + \mathbf{R}\left(\frac{\mathbf{A}}{k}\right) e^{\mathbf{B}/k}.$$

Subtracting this from $\mathbf{S}_k = \mathbf{I} + \frac{\mathbf{A} + \mathbf{B}}{k} + \mathbf{R}(\frac{\mathbf{A} + \mathbf{B}}{k})$, the constant and linear terms cancel, and the triangle inequality with $a/k, b/k, c/k \leq 1$ gives

$$\begin{aligned} \|\mathbf{S}_k - \mathbf{T}_k\|_2 &\leq \frac{c^2 e}{2k^2} + \frac{b^2 e}{2k^2} + \frac{ab}{k^2} + \frac{ab^2 e}{2k^3} + \frac{a^2 e^2}{2k^2} \\ &\leq \frac{M}{k^2}, \quad M = \frac{(c^2 + b^2)e}{2} + ab + \frac{ab^2 e}{2} + \frac{a^2 e^2}{2}, \end{aligned}$$

where the fourth term used $1/k^3 \leq 1/k^2$. The constant M depends only on $\mathbf{A}$ and $\mathbf{B}$.

Step 4. Telescoping. For any $\mathbf{M}, \mathbf{N} \in M_n(\mathbb{C})$ and any $k \geq 1$,

$$\mathbf{M}^k - \mathbf{N}^k = \sum_{j=0}^{k-1} \mathbf{M}^j (\mathbf{M} - \mathbf{N}) \mathbf{N}^{k-1-j},$$

since the j -th summand is $\mathbf{M}^{j+1} \mathbf{N}^{k-1-j} - \mathbf{M}^j \mathbf{N}^{k-j}$ and the sum telescopes. Applying this to $\mathbf{M} = \mathbf{S}_k$, $\mathbf{N} = \mathbf{T}_k$ and using Steps 1 and 2,

$$\|e^{\mathbf{A} + \mathbf{B}} - \mathbf{T}_k^k\|_2 \leq \sum_{j=0}^{k-1} e^{jc/k} \|\mathbf{S}_k - \mathbf{T}_k\|_2 e^{(k-1-j)c/k} \leq k e^c \frac{M}{k^2} = \frac{e^c M}{k}.$$

The right side tends to 0, so $\mathbf{T}_k^k \rightarrow e^{\mathbf{A} + \mathbf{B}}$ in the spectral norm, hence entrywise and in every norm by Corollary 16.2.4. This proves the formula. ■

Nothing in the proof used that the matrices are Hermitian: the formula holds for all complex matrices. What it costs is a limit, and every consequence below is obtained by pushing an inequality through that limit.

21.12.2 The trace of a product of two positive matrices

Golden–Thompson compares $\text{tr } e^{\mathbf{A} + \mathbf{B}}$ with $\text{tr}(e^{\mathbf{A}} e^{\mathbf{B}})$. Before comparing them we should know that the second number is real, which is not obvious: the product of two Hermitian matrices is almost never Hermitian. One identity settles this, and it will be used again.

Lemma 21.12.2. Powers of a Product of Two Positive Matrices

Let $\mathbf{X}, \mathbf{Y} \in M_n(\mathbb{C})$ be positive semidefinite and let $k \geq 1$. Then

$$\text{tr}((\mathbf{XY})^k) = \text{tr}((\mathbf{X}^{1/2} \mathbf{Y} \mathbf{X}^{1/2})^k),$$

and this number is real and ≥ 0 .

Proof. Let $\mathbf{X}^{1/2}$ be the positive square root of Theorem 13.2.1, and put $\mathbf{S} = \mathbf{X}^{1/2}\mathbf{Y}\mathbf{X}^{1/2}$. We claim

$$(\mathbf{XY})^k = \mathbf{X}^{1/2} \mathbf{S}^{k-1} \mathbf{X}^{1/2} \mathbf{Y} \quad (k \geq 1).$$

For $k = 1$ the right side is $\mathbf{X}^{1/2}\mathbf{X}^{1/2}\mathbf{Y} = \mathbf{XY}$. Assuming it for k , and using $\mathbf{X}^{1/2}\mathbf{Y}\mathbf{X}\mathbf{Y} = \mathbf{S}\mathbf{X}^{1/2}\mathbf{Y}$,

$$(\mathbf{XY})^{k+1} = \mathbf{X}^{1/2} \mathbf{S}^{k-1} (\mathbf{X}^{1/2} \mathbf{Y} \mathbf{X} \mathbf{Y}) = \mathbf{X}^{1/2} \mathbf{S}^k \mathbf{X}^{1/2} \mathbf{Y},$$

which is the claim for $k + 1$. Taking traces and moving the leading $\mathbf{X}^{1/2}$ to the back, which Theorem 0.9.6 (3) permits,

$$\operatorname{tr}((\mathbf{XY})^k) = \operatorname{tr}(\mathbf{S}^{k-1} \mathbf{X}^{1/2} \mathbf{Y} \mathbf{X}^{1/2}) = \operatorname{tr}(\mathbf{S}^k).$$

Now $\mathbf{S} = (\mathbf{X}^{1/2})^* \mathbf{Y} \mathbf{X}^{1/2} \succeq \mathbf{O}$ by Proposition 13.1.2 (a), since $\mathbf{X}^{1/2}$ is Hermitian. So $\mathbf{S}^k$ is Hermitian with eigenvalues the k -th powers of those of $\mathbf{S}$, all ≥ 0 by Theorem 13.1.4 (b). Its trace is the sum of its eigenvalues (Theorem 9.3.3), hence real and ≥ 0 . ■

Quick check. Why is $\operatorname{tr}(e^{\mathbf{A}} e^{\mathbf{B}})$ a real number for Hermitian $\mathbf{A}$ and $\mathbf{B}$, even though $e^{\mathbf{A}} e^{\mathbf{B}}$ is not Hermitian?

Solution. For Hermitian $\mathbf{A}$ with spectral resolution $\mathbf{A} = \sum_i \lambda_i \mathbf{P}_i$, Theorem 12.8.5 (a) gives $e^{\mathbf{A}} = \sum_i e^{\lambda_i} \mathbf{P}_i$, which is Hermitian with positive eigenvalues, hence positive definite by Theorem 13.1.4 (b). The same holds for $e^{\mathbf{B}}$. So Lemma 21.12.2 with $k = 1$ applies: $\operatorname{tr}(e^{\mathbf{A}} e^{\mathbf{B}})$ is real and ≥ 0 . The product itself is not Hermitian unless $\mathbf{A}$ and $\mathbf{B}$ commute, but its trace does not notice.

The engine of Golden–Thompson is the next inequality. It says that in a trace, the powers may be moved off the product and onto the two factors — provided the exponent is a power of two. That restriction is not an accident of the proof technique here; it is what makes a repeated Cauchy–Schwarz argument close.

Lemma 21.12.3. The Trace Power Inequality

Let $r \geq 0$ and $m = 2^r$.

- (a) $\operatorname{tr}((\mathbf{A}^*)^m \mathbf{A}^m) \leq \operatorname{tr}((\mathbf{A} \mathbf{A}^*)^m)$ for every $\mathbf{A} \in M_n(\mathbb{C})$.
- (b) $\operatorname{tr}((\mathbf{X} \mathbf{Y})^m) \leq \operatorname{tr}(\mathbf{X}^m \mathbf{Y}^m)$ for all positive semidefinite $\mathbf{X}, \mathbf{Y} \in M_n(\mathbb{C})$.

Idea. Both halves are proved together by induction on r , because each step of one needs the other at the previous stage. The only inequality used is Cauchy–Schwarz for the Frobenius inner product, in the form $|\operatorname{tr}(\mathbf{M}^2)| \leq \operatorname{tr}(\mathbf{M}^* \mathbf{M})$: a trace of a square is bounded by a trace of a product of a matrix with its adjoint. Everything else is cyclic rearrangement of words in $\mathbf{X}$ and $\mathbf{Y}$, or in $\mathbf{A}$ and $\mathbf{A}^*$. The role of (a) is to absorb the mismatch that Cauchy–Schwarz

creates: it turns $(\mathbf{A}^*)^m \mathbf{A}^m$ — the wrong grouping — into $(\mathbf{A}\mathbf{A}^*)^m$, which is a word one can feed back into (b).

Proof. Throughout, $\langle \mathbf{M}, \mathbf{N} \rangle = \text{tr}(\mathbf{N}^* \mathbf{M})$ is the Frobenius inner product of Chapter 11 §01, with induced norm $\|\mathbf{M}\|_F$.

Claim 1. $|\text{tr}(\mathbf{M}^2)| \leq \text{tr}(\mathbf{M}^* \mathbf{M})$ for every $\mathbf{M} \in M_n(\mathbb{C})$. Since $(\mathbf{M}^*)^* = \mathbf{M}$, we have $\text{tr}(\mathbf{M}^2) = \text{tr}((\mathbf{M}^*)^* \mathbf{M}) = \langle \mathbf{M}, \mathbf{M}^* \rangle$. By Theorem 11.1.5, $|\langle \mathbf{M}, \mathbf{M}^* \rangle| \leq \|\mathbf{M}\|_F \|\mathbf{M}^*\|_F$. Finally $\|\mathbf{M}^*\|_F^2 = \text{tr}(\mathbf{M}\mathbf{M}^*) = \text{tr}(\mathbf{M}^* \mathbf{M}) = \|\mathbf{M}\|_F^2$ by Theorem 0.9.6 (3), so the bound is $\|\mathbf{M}\|_F^2 = \text{tr}(\mathbf{M}^* \mathbf{M})$.

Claim 2. For positive semidefinite $\mathbf{X}, \mathbf{Y}$ and $q \geq 1$, $\text{tr}((\mathbf{X}\mathbf{Y}^2\mathbf{X})^q) = \text{tr}((\mathbf{X}^2\mathbf{Y}^2)^q)$; and for any $\mathbf{A}$ and $q \geq 1$, $\text{tr}((\mathbf{A}^2(\mathbf{A}^*)^2)^q) = \text{tr}((\mathbf{P}\mathbf{Q})^q)$, where $\mathbf{P} = \mathbf{A}\mathbf{A}^*$ and $\mathbf{Q} = \mathbf{A}^*\mathbf{A}$. Both are cyclic rearrangements, justified by Theorem 0.9.6 (3). Induction on q gives $(\mathbf{X}\mathbf{Y}^2\mathbf{X})^q = \mathbf{X}(\mathbf{Y}^2\mathbf{X}^2)^{q-1}\mathbf{Y}^2\mathbf{X}$; moving the leading $\mathbf{X}$ to the back turns its trace into $\text{tr}((\mathbf{Y}^2\mathbf{X}^2)^q)$, which is $\text{tr}((\mathbf{X}^2\mathbf{Y}^2)^q)$ by one more cyclic shift. For the second identity, $\mathbf{A}^2(\mathbf{A}^*)^2 = \mathbf{A}\mathbf{P}\mathbf{A}^*$, and induction on q gives $(\mathbf{A}\mathbf{P}\mathbf{A}^*)^q = \mathbf{A}\mathbf{P}(\mathbf{Q}\mathbf{P})^{q-1}\mathbf{A}^*$; moving the leading $\mathbf{A}$ to the back turns its trace into $\text{tr}(\mathbf{P}(\mathbf{Q}\mathbf{P})^{q-1}\mathbf{Q}) = \text{tr}((\mathbf{P}\mathbf{Q})^q)$.

We now induct on r , proving (a) and (b) simultaneously.

Base $r = 0$. Part (a) reads $\text{tr}(\mathbf{A}^*\mathbf{A}) \leq \text{tr}(\mathbf{A}\mathbf{A}^*)$, an equality by Theorem 0.9.6 (3). Part (b) reads $\text{tr}(\mathbf{X}\mathbf{Y}) \leq \text{tr}(\mathbf{Y}\mathbf{X})$.

Step. Assume (a) and (b) for r , and write $q = 2^r$.

Part (b) for $2q$. Put $\mathbf{A} = \mathbf{X}\mathbf{Y}$, so $\mathbf{A}^* = \mathbf{Y}\mathbf{X}$ because $\mathbf{X}$ and $\mathbf{Y}$ are Hermitian, and put $\mathbf{M} = \mathbf{A}^q = (\mathbf{X}\mathbf{Y})^q$. Then $\mathbf{M}^2 = (\mathbf{X}\mathbf{Y})^{2q}$ and $\mathbf{M}^* \mathbf{M} = (\mathbf{A}^*)^q \mathbf{A}^q$. The number $\text{tr}(\mathbf{M}^2) = \text{tr}((\mathbf{X}\mathbf{Y})^{2q})$ is real and ≥ 0 by Lemma 21.12.2, so it equals its own absolute value, and Claim 1 gives

$$\text{tr}((\mathbf{X}\mathbf{Y})^{2q}) \leq |\text{tr}(\mathbf{M}^2)| \leq \text{tr}((\mathbf{A}^*)^q \mathbf{A}^q).$$

By the inductive hypothesis (a),

$$\text{tr}((\mathbf{A}^*)^q \mathbf{A}^q) \leq \text{tr}((\mathbf{A}\mathbf{A}^*)^q) = \text{tr}((\mathbf{X}\mathbf{Y}^2\mathbf{X})^q) = \text{tr}((\mathbf{X}^2\mathbf{Y}^2)^q),$$

the last step by Claim 2. Now $\mathbf{X}^2 = \mathbf{X}^* \mathbf{X} \succeq \mathbf{O}$ and $\mathbf{Y}^2 \succeq \mathbf{O}$ by Theorem 13.1.4 (c), so the inductive hypothesis (b) applies to the pair $\mathbf{X}^2, \mathbf{Y}^2$ and gives

$$\text{tr}((\mathbf{X}^2\mathbf{Y}^2)^q) \leq \text{tr}(\mathbf{X}^{2q}\mathbf{Y}^{2q}).$$

Chaining the three displays proves (b) for $2q$.

Part (a) for $2q$. Put $\mathbf{B} = \mathbf{A}^2$, so $\mathbf{B}^* = (\mathbf{A}^*)^2$, $\mathbf{A}^{2q} = \mathbf{B}^q$, and $(\mathbf{A}^*)^{2q} = (\mathbf{B}^*)^q$. The inductive hypothesis (a), applied to $\mathbf{B}$, and then Claim 2 give

$$\text{tr}((\mathbf{A}^*)^{2q} \mathbf{A}^{2q}) = \text{tr}((\mathbf{B}^*)^q \mathbf{B}^q) \leq \text{tr}((\mathbf{B}\mathbf{B}^*)^q) = \text{tr}((\mathbf{P}\mathbf{Q})^q),$$

with $\mathbf{P} = \mathbf{A}\mathbf{A}^*$ and $\mathbf{Q} = \mathbf{A}^*\mathbf{A}$, both positive semidefinite by Theorem 13.1.4 (c). The inductive hypothesis (b) applied to $\mathbf{P}, \mathbf{Q}$ gives $\text{tr}((\mathbf{P}\mathbf{Q})^q) \leq \text{tr}(\mathbf{P}^q \mathbf{Q}^q)$. Since $\mathbf{P}^q$ and $\mathbf{Q}^q$ are Hermitian, $\text{tr}(\mathbf{P}^q \mathbf{Q}^q) = \langle \mathbf{P}^q, \mathbf{Q}^q \rangle$, and this number is real and ≥ 0 by Lemma 21.12.2, both matrices being positive semidefinite. So Theorem 11.1.5 gives

$$\text{tr}(\mathbf{P}^q \mathbf{Q}^q) = \langle \mathbf{P}^q, \mathbf{Q}^q \rangle \leq \|\mathbf{P}^q\|_F \|\mathbf{Q}^q\|_F = \text{tr}(\mathbf{P}^{2q})^{1/2} \text{tr}(\mathbf{Q}^{2q})^{1/2}.$$

Write $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*$ for a singular value decomposition (Theorem 13.9.1), with $\mathbf{\Sigma} = \text{diag}(\sigma_1, \dots, \sigma_n)$. Then $\mathbf{P} = \mathbf{U}\mathbf{\Sigma}^2\mathbf{U}^*$ and $\mathbf{Q} = \mathbf{V}\mathbf{\Sigma}^2\mathbf{V}^*$, so $\mathbf{P}$ and $\mathbf{Q}$ are both unitarily similar to $\mathbf{\Sigma}^2$ and therefore have the same eigenvalues with multiplicity. Hence $\text{tr}(\mathbf{P}^{2q}) = \text{tr}(\mathbf{Q}^{2q})$, the displayed bound is $\text{tr}(\mathbf{P}^{2q})$, and

$$\text{tr}((\mathbf{A}^*)^{2q}\mathbf{A}^{2q}) \leq \text{tr}(\mathbf{P}^{2q}) = \text{tr}((\mathbf{A}\mathbf{A}^*)^{2q}).$$

This is (a) for $2q$, and the induction is complete. ■

The restriction to $m = 2^r$ is a restriction of this proof, not of the truth: (b) holds for every integer $m \geq 1$. It even holds for every real $m \geq 1$, once $(\mathbf{X}\mathbf{Y})^m$ is read as $(\mathbf{X}^{1/2}\mathbf{Y}\mathbf{X}^{1/2})^m$, which Lemma 21.12.2 says is the same thing for integer exponents and which is the only reading available for the others, since $\mathbf{X}\mathbf{Y}$ is neither Hermitian nor positive semidefinite. The general case is a theorem of Lieb and Thirring and is not proved here; nothing below needs it, because a limit along the powers of two is all Lie’s formula asks for.

21.12.3 The inequality of Golden and Thompson

Now the two halves fit together. Lie’s formula writes $e^{\mathbf{A}+\mathbf{B}}$ as a limit of m -th powers of $e^{\mathbf{A}/m}e^{\mathbf{B}/m}$, and the trace power inequality moves the exponent m off the product and onto the two factors, where it reassembles $e^{\mathbf{A}}$ and $e^{\mathbf{B}}$.

Theorem 21.12.4. The Golden–Thompson Inequality

Let $\mathbf{A}, \mathbf{B} \in M_n(\mathbb{C})$ be Hermitian. Then

$$\text{tr} e^{\mathbf{A}+\mathbf{B}} \leq \text{tr}(e^{\mathbf{A}}e^{\mathbf{B}}).$$

Both sides are real and positive, and equality holds if $\mathbf{AB} = \mathbf{BA}$.

Idea. Run the trace power inequality at the m -th root of each exponential and let m run through the powers of two. At each m the right-hand side is exactly $\text{tr}(e^{\mathbf{A}}e^{\mathbf{B}})$, independent of m , while the left-hand side converges to $\text{tr} e^{\mathbf{A}+\mathbf{B}}$ by Lie’s formula. A nonstrict inequality survives a limit.

Proof. Fix $r \geq 0$ and put $m = 2^r$, $\mathbf{X} = e^{\mathbf{A}/m}$, and $\mathbf{Y} = e^{\mathbf{B}/m}$. Since $\mathbf{A}/m$ is Hermitian with spectral resolution $\sum_i \mu_i \mathbf{P}_i$, Theorem 12.8.5 (a) gives $\mathbf{X} = \sum_i e^{\mu_i} \mathbf{P}_i$, which is Hermitian with all eigenvalues $e^{\mu_i} > 0$; so $\mathbf{X} \succ \mathbf{O}$ by Theorem 13.1.5 (b), and likewise $\mathbf{Y} \succ \mathbf{O}$. As in Step 1 of the proof of Lemma 21.12.1, $\mathbf{X}^m = e^{\mathbf{A}}$ and $\mathbf{Y}^m = e^{\mathbf{B}}$.

By Lemma 21.12.3 (b),

$$\text{tr}\left((e^{\mathbf{A}/m}e^{\mathbf{B}/m})^m\right) \leq \text{tr}(\mathbf{X}^m\mathbf{Y}^m) = \text{tr}(e^{\mathbf{A}}e^{\mathbf{B}}).$$

The right side does not depend on r . By Lemma 21.12.1 the matrices $(e^{\mathbf{A}/m}e^{\mathbf{B}/m})^m$ converge entrywise to $e^{\mathbf{A}+\mathbf{B}}$ as $m = 2^r \rightarrow \infty$, so their traces, being sums of n entries, converge to $\text{tr} e^{\mathbf{A}+\mathbf{B}}$. A nonstrict inequality survives a limit, so $\text{tr} e^{\mathbf{A}+\mathbf{B}} \leq \text{tr}(e^{\mathbf{A}}e^{\mathbf{B}})$.

Both sides are real and ≥ 0 : the left because $\mathbf{A} + \mathbf{B}$ is Hermitian, so $e^{\mathbf{A}+\mathbf{B}} \succ \mathbf{O}$ by the argument just given, and the right by Lemma 21.12.2 with $k = 1$. They are in fact positive, since a positive definite matrix has positive trace. If $\mathbf{AB} = \mathbf{BA}$, then $e^{\mathbf{A}}e^{\mathbf{B}} = e^{\mathbf{A}+\mathbf{B}}$ by Theorem 10.9.4 (d), and the two sides are equal. ■

A two-by-two family shows that the inequality is genuinely an inequality, and lets the reader see what the two sides look like.

Example. The inequality on a two-by-two family

For real a, b let

$$\mathbf{A} = \begin{pmatrix} a & 0 \\ 0 & -a \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 0 & b \\ b & 0 \end{pmatrix}.$$

Compute both sides of Golden–Thompson, and decide when the inequality is strict.

Solution. Both matrices are real symmetric. Write $\mathbf{M} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, so that $\mathbf{B} = b\mathbf{M}$ and $\mathbf{M}^2 = \mathbf{I}$. Then $\mathbf{B}^{2m} = b^{2m}\mathbf{I}$ and $\mathbf{B}^{2m+1} = b^{2m+1}\mathbf{M}$, so splitting the exponential series into even and odd terms gives

$$e^{\mathbf{B}} = \cosh(b)\mathbf{I} + \sinh(b)\mathbf{M} = \begin{pmatrix} \cosh b & \sinh b \\ \sinh b & \cosh b \end{pmatrix},$$

for every real b . Since $e^{\mathbf{A}} = \text{diag}(e^a, e^{-a})$, the diagonal entries of $e^{\mathbf{A}}e^{\mathbf{B}}$ are $e^a \cosh b$ and $e^{-a} \cosh b$, whence

$$\text{tr}(e^{\mathbf{A}}e^{\mathbf{B}}) = (e^a + e^{-a}) \cosh b = 2 \cosh a \cosh b.$$

For the left side, $\mathbf{A} + \mathbf{B} = \begin{pmatrix} a & b \\ b & -a \end{pmatrix}$ has trace 0 and determinant $-(a^2 + b^2)$, so its eigenvalues are $\pm d$ with $d = \sqrt{a^2 + b^2}$. By Theorem 12.8.5 (a) the eigenvalues of $e^{\mathbf{A}+\mathbf{B}}$ are e^d and e^{-d} , so $\text{tr } e^{\mathbf{A}+\mathbf{B}} = 2 \cosh d$ by Theorem 9.3.3. The inequality therefore reads

$$\cosh \sqrt{a^2 + b^2} \leq \cosh a \cosh b,$$

a statement about numbers, and it can be checked term by term. Expanding both sides as power series with nonnegative terms, the coefficient of $a^{2i}b^{2j}$ is $\binom{i+j}{i}/(2i+2j)!$ on the left and $1/((2i)!(2j)!)$ on the right, so what is needed is $\binom{i+j}{i} \leq \binom{2i+2j}{2i}$. That holds because doubling each of i chosen items out of $i+j$ produces $2i$ chosen items out of $2i+2j$, injectively; and it is strict as soon as $i, j \geq 1$, since then some $2i$ -subset splits a doubled pair and so is not in the image.

Equality therefore forces every term with $i, j \geq 1$ to vanish, that is $a = 0$ or $b = 0$ — exactly the cases where $\mathbf{A}$ and $\mathbf{B}$ commute. For $a = b = 1$ the two sides are $2 \cosh \sqrt{2} \approx 4.3564$ and $2 \cosh^2 1 \approx 4.7622$.

Warning. Golden–Thompson is a statement about traces and about nothing else. It is tempting to guess the Loewner-order version $e^{\mathbf{A}+\mathbf{B}} \preceq e^{\mathbf{A}/2}e^{\mathbf{B}}e^{\mathbf{A}/2}$, whose right side is Hermitian and positive semidefinite and has the same trace as $e^{\mathbf{A}}e^{\mathbf{B}}$. It is false in general. Indeed both sides have the same determinant, namely $e^{\text{tr } \mathbf{A} + \text{tr } \mathbf{B}}$ by Theorem 10.9.4 (e) —

and if $\mathbf{M} \succeq \mathbf{N} \succ \mathbf{O}$ with $\det \mathbf{M} = \det \mathbf{N}$, then $\mathbf{Z} = \mathbf{N}^{-1/2} \mathbf{M} \mathbf{N}^{-1/2} \succeq \mathbf{I}$ by Theorem 13.5.3 (a), so all eigenvalues of $\mathbf{Z}$ are ≥ 1 while their product is $\det \mathbf{Z} = \det \mathbf{M} / \det \mathbf{N} = 1$; hence they are all 1, the Hermitian matrix $\mathbf{Z}$ equals $\mathbf{I}$, and $\mathbf{M} = \mathbf{N}^{1/2} \mathbf{Z} \mathbf{N}^{1/2} = \mathbf{N}$. So the Loewner inequality would force $e^{\mathbf{A}+\mathbf{B}} = e^{\mathbf{A}/2} e^{\mathbf{B}} e^{\mathbf{A}/2}$, and taking traces would force equality in Golden–Thompson, which [Example \(The inequality on a two-by-two family\)](#) refutes for $a = b = 1$.

21.12.4 Three matrices are one too many

Golden–Thompson looks as though it should iterate. It does not. The three-matrix statement

$$\operatorname{tr} e^{\mathbf{A}+\mathbf{B}+\mathbf{C}} \leq \operatorname{tr}(e^{\mathbf{A}} e^{\mathbf{B}} e^{\mathbf{C}})$$

is **false**, and the failure is not marginal: the right side can be negative while the left side, a trace of a positive definite matrix, is positive. A family of two-by-two real symmetric matrices shows it, exactly.

Example. The three-matrix version fails

Let $t > 0$ and let $\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3$ be the orthogonal projections of $\mathbb{R}^2$ onto the three lines through the origin at angles $0, 2\pi/3$, and $4\pi/3$:

$$\mathbf{P}_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad \mathbf{P}_2 = \frac{1}{4} \begin{pmatrix} 1 & -\sqrt{3} \\ -\sqrt{3} & 3 \end{pmatrix}, \quad \mathbf{P}_3 = \frac{1}{4} \begin{pmatrix} 1 & \sqrt{3} \\ \sqrt{3} & 3 \end{pmatrix}.$$

Put $\mathbf{A} = t\mathbf{P}_1$, $\mathbf{B} = t\mathbf{P}_2$, $\mathbf{C} = t\mathbf{P}_3$, all symmetric. Compute both sides of the three-matrix inequality and show that it fails.

Solution. Each $\mathbf{P}_i$ is $\mathbf{u}_i \mathbf{u}_i^\top$ for a unit vector $\mathbf{u}_i$, with $\mathbf{u}_1 = (1, 0)$, $\mathbf{u}_2 = (-\frac{1}{2}, \frac{\sqrt{3}}{2})$, $\mathbf{u}_3 = (-\frac{1}{2}, -\frac{\sqrt{3}}{2})$; each is symmetric with $\mathbf{P}_i^2 = \mathbf{P}_i$. Adding the three matrices gives $\mathbf{P}_1 + \mathbf{P}_2 + \mathbf{P}_3 = \frac{3}{2}\mathbf{I}$, so $\mathbf{A} + \mathbf{B} + \mathbf{C} = \frac{3t}{2}\mathbf{I}$ and

$$\operatorname{tr} e^{\mathbf{A}+\mathbf{B}+\mathbf{C}} = 2e^{3t/2}.$$

For the right side, $\mathbf{P}_i^2 = \mathbf{P}_i$ makes the exponential series collapse: $e^{t\mathbf{P}} = \mathbf{I} + \alpha\mathbf{P}$ with $\alpha = e^t - 1$, because $\sum_{j \geq 1} t^j \mathbf{P}^j / j! = (\sum_{j \geq 1} t^j / j!) \mathbf{P}$. Expanding the product of three such factors and using $\operatorname{tr} \mathbf{I} = 2$, $\operatorname{tr} \mathbf{P}_i = 1$, $\operatorname{tr}(\mathbf{P}_i \mathbf{P}_j) = (\mathbf{u}_i \cdot \mathbf{u}_j)^2$, and $\operatorname{tr}(\mathbf{P}_1 \mathbf{P}_2 \mathbf{P}_3) = (\mathbf{u}_1 \cdot \mathbf{u}_2)(\mathbf{u}_2 \cdot \mathbf{u}_3)(\mathbf{u}_3 \cdot \mathbf{u}_1)$,

$$\operatorname{tr}(e^{\mathbf{A}} e^{\mathbf{B}} e^{\mathbf{C}}) = 2 + 3\alpha + \frac{3}{4}\alpha^2 - \frac{1}{8}\alpha^3,$$

since each pairwise dot product is $-\frac{1}{2}$, so each $\operatorname{tr}(\mathbf{P}_i \mathbf{P}_j) = \frac{1}{4}$ and the triple product is $-\frac{1}{8}$. In terms of t this is

$$\operatorname{tr}(e^{\mathbf{A}} e^{\mathbf{B}} e^{\mathbf{C}}) = \frac{1}{8}(-e^{3t} + 9e^{2t} + 9e^t - 1).$$

The cubic term is negative, and for large t it wins. Precisely, if $e^t \geq 10$ then $e^{2t} \geq 10e^t$ and $e^{3t} \geq 10e^{2t} = 9e^{2t} + e^{2t} \geq 9e^{2t} + 10e^t$, so the bracket is negative and $\operatorname{tr}(e^{\mathbf{A}} e^{\mathbf{B}} e^{\mathbf{C}}) < 0$, while $\operatorname{tr} e^{\mathbf{A}+\mathbf{B}+\mathbf{C}} = 2e^{3t/2} > 0$. Taking $t = 3$, where $e^3 > 20$, already does this: the two sides are $2e^{4.5} \approx 180.0$ and $\frac{1}{8}(-e^9 + 9e^6 + 9e^3 - 1) \approx -536.6$.

The failure is not confined to large t . At $t = 1$ the two sides are $2e^{3/2} \approx 8.9634$ and

$\frac{1}{8}(-e^3 + 9e^2 + 9e - 1) \approx 8.7351$, and at $t = 2$ they are $2e^3 \approx 40.171$ and $\frac{1}{8}(-e^6 + 9e^4 + 9e^2 - 1) \approx 19.182$.

The numbers in the example are exact, and the two closed forms may be compared by hand; only their decimal expansions are approximations.

Remark. There is a second, more basic obstruction. For three Hermitian matrices the number $\text{tr}(e^{\mathbf{A}}e^{\mathbf{B}}e^{\mathbf{C}})$ need not be **real**, so no inequality can even be written down. With

$$\mathbf{A} = a \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \mathbf{B} = b \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \mathbf{C} = c \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

each of the three squares is a multiple of $\mathbf{I}$, so each exponential splits into a hyperbolic cosine times $\mathbf{I}$ plus a hyperbolic sine times the matrix, and multiplying out gives

$$\text{tr}(e^{\mathbf{A}}e^{\mathbf{B}}e^{\mathbf{C}}) = 2 \cosh a \cosh b \cosh c + 2i \sinh a \sinh b \sinh c,$$

because the product of the three matrices above is $i\mathbf{I}$ while every shorter product among them has trace 0. The imaginary part is nonzero whenever a, b, c are all nonzero. With two matrices this cannot happen, by [Lemma 21.12.2](#).

Remark. Two of the three matrices commuting is enough to restore the inequality, and exercise [Exercise \(C1: Two out of three is enough\)](#) asks for the three cases. So the example above is as economical as it can be: no two of $\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3$ commute.

21.12.5 Klein's inequality and relative entropy

Golden–Thompson is one of a family of trace inequalities whose proofs all end in the same place: a scalar convexity inequality summed against nonnegative weights that add to one. The next theorem is the cleanest member of the family. It needs one fact about convex functions of a real variable, which Chapter 18 recorded only in an exercise, so we prove it here.

Lemma 21.12.5. A Convex Function Lies Above Its Tangent

Let $I \subseteq \mathbb{R}$ be an open interval and let $f: I \rightarrow \mathbb{R}$ be convex and differentiable. Then

$$f(s) \geq f(t) + f'(t)(s - t) \quad \text{for all } s, t \in I.$$

Proof. Fix $s, t \in I$. If $s = t$ both sides are $f(t)$. Assume $s \neq t$ and let $0 < \theta \leq 1$. The point $t + \theta(s - t) = (1 - \theta)t + \theta s$ lies in I , and Definition 18.10.1 gives

$$f(t + \theta(s - t)) \leq (1 - \theta)f(t) + \theta f(s),$$

that is, $(f(t + \theta(s - t)) - f(t))/\theta \leq f(s) - f(t)$. Writing $h = \theta(s - t)$, which is nonzero and tends to 0 as $\theta \rightarrow 0^+$, the left side equals

$$(s - t) \cdot \frac{f(t + h) - f(t)}{h} \longrightarrow (s - t)f'(t),$$

by differentiability of f at t . A nonstrict inequality survives a limit, so $(s - t)f'(t) \leq f(s) - f(t)$, which is the claim. ■

Theorem 21.12.6. Klein's Inequality

Let $I \subseteq \mathbb{R}$ be an open interval and let $f: I \rightarrow \mathbb{R}$ be convex and differentiable. Let $\mathbf{A}, \mathbf{B} \in M_n(\mathbb{C})$ be Hermitian with $\text{spec}(\mathbf{A}) \subseteq I$ and $\text{spec}(\mathbf{B}) \subseteq I$. Then

$$\text{tr}(f(\mathbf{A}) - f(\mathbf{B}) - (\mathbf{A} - \mathbf{B})f'(\mathbf{B})) \geq 0,$$

where $f(\mathbf{A})$, $f(\mathbf{B})$, and $f'(\mathbf{B})$ are the functional calculus of Definition 12.8.2.

Idea. Compute the trace in an orthonormal eigenbasis of $\mathbf{B}$. Each diagonal entry then involves $\mathbf{B}$ only through one eigenvalue μ_j , and involves $\mathbf{A}$ only through the numbers $c_{ij} = |\langle \mathbf{v}_j, \mathbf{u}_i \rangle|^2$, the overlaps between the two eigenbases. Those numbers are nonnegative and each column of the array sums to 1, which is exactly what is needed to write the whole trace as a weighted average of the scalar quantities $f(\lambda_i) - f(\mu_j) - f'(\mu_j)(\lambda_i - \mu_j)$. Every one of those is ≥ 0 by Lemma 21.12.5.

Proof. By Corollary 12.4.3 there are orthonormal bases $(\mathbf{u}_1, \dots, \mathbf{u}_n)$ and $(\mathbf{v}_1, \dots, \mathbf{v}_n)$ of $\mathbb{C}^n$ with $\mathbf{A}\mathbf{u}_i = \lambda_i\mathbf{u}_i$ and $\mathbf{B}\mathbf{v}_j = \mu_j\mathbf{v}_j$; all λ_i and μ_j lie in I . Put $c_{ij} = |\langle \mathbf{v}_j, \mathbf{u}_i \rangle|^2 \geq 0$. Expanding $\mathbf{v}_j$ in the orthonormal basis $(\mathbf{u}_i)$ gives $\sum_i c_{ij} = \|\mathbf{v}_j\|^2 = 1$ for each j .

For any $\mathbf{M} \in M_n(\mathbb{C})$ we have $\text{tr} \mathbf{M} = \sum_j \langle \mathbf{M}\mathbf{v}_j, \mathbf{v}_j \rangle$: with $\mathbf{V}$ the unitary matrix whose columns are the $\mathbf{v}_j$, the sum is $\text{tr}(\mathbf{V}^*\mathbf{M}\mathbf{V}) = \text{tr}(\mathbf{M}\mathbf{V}\mathbf{V}^*) = \text{tr} \mathbf{M}$ by Theorem 0.9.6 (3). We evaluate the three terms at $\mathbf{v}_j$.

By Definition 12.8.2, $f(\mathbf{A})\mathbf{u}_i = f(\lambda_i)\mathbf{u}_i$, so expanding $\mathbf{v}_j = \sum_i \langle \mathbf{v}_j, \mathbf{u}_i \rangle \mathbf{u}_i$ and using orthonormality,

$$\langle f(\mathbf{A})\mathbf{v}_j, \mathbf{v}_j \rangle = \sum_i f(\lambda_i) c_{ij}, \quad \langle \mathbf{A}\mathbf{v}_j, \mathbf{v}_j \rangle = \sum_i \lambda_i c_{ij}.$$

Also $f(\mathbf{B})\mathbf{v}_j = f(\mu_j)\mathbf{v}_j$ and $f'(\mathbf{B})\mathbf{v}_j = f'(\mu_j)\mathbf{v}_j$, so $\langle f(\mathbf{B})\mathbf{v}_j, \mathbf{v}_j \rangle = f(\mu_j)$ and

$$\langle (\mathbf{A} - \mathbf{B})f'(\mathbf{B})\mathbf{v}_j, \mathbf{v}_j \rangle = f'(\mu_j) \left(\sum_i \lambda_i c_{ij} - \mu_j \right).$$

Summing over j , and using $\sum_i c_{ij} = 1$ to write $f(\mu_j) = \sum_i c_{ij} f(\mu_j)$ and $\mu_j = \sum_i c_{ij} \mu_j$,

$$\text{tr}(f(\mathbf{A}) - f(\mathbf{B}) - (\mathbf{A} - \mathbf{B})f'(\mathbf{B})) = \sum_{i,j} c_{ij} [f(\lambda_i) - f(\mu_j) - f'(\mu_j)(\lambda_i - \mu_j)].$$

Every bracket is ≥ 0 by Lemma 21.12.5, and every $c_{ij} \geq 0$. This proves the inequality. ■

The corollary that made Klein's inequality famous concerns density matrices, the positive semidefinite matrices of trace 1 that Chapter 18 §08 met as a convex body (Definition 18.8.3).

It needs the derivative of the logarithm, which the book has not yet computed.

Lemma 21.12.7. The Derivative of the Logarithm

The function $\log: (0, \infty) \rightarrow \mathbb{R}$ of Lemma 18.11.1 is differentiable with $\log'(y) = 1/y$. Consequently $g(t) = t \log t$ is differentiable on $(0, \infty)$ with $g'(t) = \log t + 1$ and $g''(t) = 1/t$, and g is strictly convex there.

Proof. Continuity. By Lemma 18.11.1 (a) and (b), $\exp: \mathbb{R} \rightarrow (0, \infty)$ is a strictly increasing bijection, so $\log$ is strictly increasing. Fix $y > 0$, put $s = \log y$, and let $\varepsilon > 0$. Then $e^{s-\varepsilon} < y < e^{s+\varepsilon}$, and for y' in that open interval, applying the increasing $\log$ gives $s - \varepsilon < \log y' < s + \varepsilon$. So $\log$ is continuous at y .

Differentiability. Let $h \neq 0$ be small enough that $y + h > 0$, and put $u = \log(y + h) - s$, so that $y + h = e^{s+u} = ye^u$ and $h = y(e^u - 1)$. Since $\log$ is injective, $u \neq 0$, and $u \rightarrow 0$ as $h \rightarrow 0$ by continuity. Hence

$$\frac{\log(y + h) - \log y}{h} = \frac{u}{y(e^u - 1)} \rightarrow \frac{1}{y},$$

because $(e^u - 1)/u \rightarrow 1$: the exponential is its own derivative (Theorem 10.9.4 (b) with $n = 1$), and its value at 0 is 1.

The function $t \log t$. For $t > 0$ and small $h \neq 0$,

$$\frac{(t + h) \log(t + h) - t \log t}{h} = t \cdot \frac{\log(t + h) - \log t}{h} + \log(t + h),$$

which tends to $t \cdot (1/t) + \log t = 1 + \log t$, using continuity of $\log$. Differentiating once more, $g''(t) = 1/t > 0$, so g is strictly convex by Lemma 18.10.5. ■

Corollary 21.12.8. Nonnegativity of the Relative Entropy

Let $\mathbf{A}, \mathbf{B} \in M_n(\mathbb{C})$ be positive definite with $\text{tr } \mathbf{A} = \text{tr } \mathbf{B} = 1$. Then

$$\text{tr}(\mathbf{A}(\log \mathbf{A} - \log \mathbf{B})) \geq 0.$$

Proof. Apply Theorem 21.12.6 on $I = (0, \infty)$ to $f(t) = t \log t$, which is convex and differentiable with $f'(t) = \log t + 1$ by Lemma 21.12.7. Both spectra lie in I by Theorem 13.1.5 (b). By Theorem 12.8.3 (b), $f'(\mathbf{B}) = \log \mathbf{B} + \mathbf{I}$. So Klein's inequality reads

$$\text{tr}(\mathbf{A} \log \mathbf{A}) - \text{tr}(\mathbf{B} \log \mathbf{B}) - \text{tr}((\mathbf{A} - \mathbf{B})(\log \mathbf{B} + \mathbf{I})) \geq 0.$$

Expanding the last trace by linearity, it equals $\text{tr}(\mathbf{A} \log \mathbf{B}) - \text{tr}(\mathbf{B} \log \mathbf{B}) + \text{tr } \mathbf{A} - \text{tr } \mathbf{B}$. Substituting and canceling $\text{tr}(\mathbf{B} \log \mathbf{B})$,

$$\text{tr}(\mathbf{A} \log \mathbf{A}) - \text{tr}(\mathbf{A} \log \mathbf{B}) - \text{tr } \mathbf{A} + \text{tr } \mathbf{B} \geq 0,$$

and $\text{tr } \mathbf{A} = \text{tr } \mathbf{B} = 1$ removes the last two terms. ■

The quantity $\operatorname{tr}(\mathbf{A}(\log \mathbf{A} - \log \mathbf{B}))$ is the **relative entropy** of $\mathbf{A}$ with respect to $\mathbf{B}$, and the corollary says it is a genuine measure of discrepancy: never negative, and zero when $\mathbf{A} = \mathbf{B}$. For diagonal $\mathbf{A}$ and $\mathbf{B}$ it is the familiar sum $\sum_i a_i(\log a_i - \log b_i)$ over two probability vectors, so the corollary contains the scalar statement as its commuting case.

Remark. Klein’s inequality is the trace-level shadow of §02’s Lemma 21.2.7 and Theorem 21.2.8: all three say that a convex function applied to a Hermitian matrix behaves, after a trace, like a convex function applied to a weighted average of eigenvalues, the weights being the overlaps between two eigenbases. The array (c_{ij}) in the proof above is doubly stochastic (Definition 18.8.4), and Chapter 20 §05 used the same array for Hoffman–Wielandt. It is the same object doing the same job.

21.12.6 What is stated here and not proved

One theorem belongs in this section by subject and is beyond its means.

Lieb’s concavity theorem. *Let $\mathbf{X} \in M_n(\mathbb{C})$ be fixed and let $p, q \geq 0$ with $p + q \leq 1$. Then the map*

$$(\mathbf{A}, \mathbf{B}) \longmapsto \operatorname{tr}(\mathbf{X}^* \mathbf{A}^p \mathbf{X} \mathbf{B}^q)$$

is jointly concave on pairs of positive definite matrices. This is a theorem of Lieb, and it is **not proved in this book**. What it needs is machinery this chapter does not build: either the theory of tensor products of positive maps, or a differentiation argument through the integral representations of §10 that would have to control two variables at once. It is recorded here because it is the natural strengthening of everything above — Golden–Thompson and the concavity of $(\mathbf{A}, \mathbf{B}) \mapsto \operatorname{tr} e^{\log \mathbf{A} + \log \mathbf{B}}$ both follow from it — and because a reader who meets it elsewhere should know where it sits.

Nothing in this book depends on it. No statement in Chapters 0 to 21 is proved using Lieb’s theorem, and no later chapter is promised a consequence of it.

21.12.7 Summary and transfer

Five earlier chapters left inequalities unproved, each of them an inequality that was supposed to hold for **every** unitarily invariant norm, or for **every** convex function. This chapter proved them, and the reason they could all be proved at once is that they are all the same statement about vectors: a majorization.

- **Majorization is a doubly stochastic map.** §01 proved the theorem of Hardy, Littlewood, and Pólya, Theorem 21.1.3, that $\mathbf{x} \prec \mathbf{y}$ exactly when $\mathbf{x} = \mathbf{D}\mathbf{y}$ for a doubly stochastic $\mathbf{D}$, and with Birkhoff’s theorem from Chapter 19, exactly when $\mathbf{x}$ lies in the convex hull of the permutations of $\mathbf{y}$ (Corollary 21.1.4).
- **Hence every convex function.** §02’s Theorem 21.2.1 turned that into $\sum_i \phi(x_i) \leq \sum_i \phi(y_i)$, and Theorem 21.2.8 turned it into the convexity of $\mathbf{A} \mapsto \operatorname{tr} f(\mathbf{A})$.
- **Hence every unitarily invariant norm.** §03 identified the symmetric gauge functions and showed that they are monotone under weak majorization (Theorem 21.3.4); §04’s Theorem 21.4.3 matched them one for one with the unitarily invariant norms of Definition 21.4.1, through $\|\mathbf{A}\| = \Phi(\sigma(\mathbf{A}))$; §05 proved von Neumann’s trace inequality; and §06’s Theorem 21.6.1 closed the circle, making $\sigma(\mathbf{A}) \prec_w \sigma(\mathbf{B})$ *equivalent* to $\|\mathbf{A}\| \leq \|\mathbf{B}\|$ for every such norm.
- **The applications.** §06 deduced Eckart–Young (Corollary 21.6.2) and the nearest-unitary

theorem (Corollary 21.6.4) in every unitarily invariant norm; §07 did the same for Lidskii (Theorem 21.7.2) and Mirsky (Theorem 21.7.4).

- **The order-theoretic half.** §08 studied operator convexity and proved the Hansen–Pedersen–Jensen inequality (Theorem 21.8.5); §09 proved that operator monotonicity of order n is exactly positivity of every $n \times n$ Loewner matrix (Theorem 21.9.5); §10 assembled Loewner’s theorem (Theorem 21.10.4), and with it Corollary 21.10.8 and Corollary 21.10.9; §11 built the matrix geometric mean (Definition 21.11.1).
- **The exponential.** This section proved Lie’s product formula, the trace power inequality, Golden–Thompson, and Klein’s inequality, and showed that three matrices break the pattern.

Six moves did the work of the chapter.

- **A majorization is a doubly stochastic map.** To prove something for all $\mathbf{x} \prec \mathbf{y}$, write $\mathbf{x} = \mathbf{D}\mathbf{y}$ and let convexity act one row at a time (§01, §02). *Transfer:* when a hypothesis is a family of inequalities between partial sums, look for the linear map that realizes it; the inequalities are usually the vertices of a polytope in disguise.
- **A symmetric gauge turns a weak majorization into a norm inequality.** A norm that cannot tell coordinates apart, and cannot see their signs, is monotone along $\prec_w$ (§03, §04). *Transfer:* to prove one inequality for a whole family of norms, find the single combinatorial statement about the underlying numbers that all of them respect.
- **A Hermitian dilation turns singular values into eigenvalues.** Chapter 17 §09’s $\mathcal{H}(\mathbf{A})$ has eigenvalues $\pm\sigma_i(\mathbf{A})$, so every theorem about eigenvalues of Hermitian matrices has a singular-value twin (§04, §07). *Transfer:* when a result is known for Hermitian matrices and wanted for all matrices, look for the Hermitian matrix whose spectrum encodes the data.
- **The Schur product theorem turns a Loewner matrix into monotonicity.** The derivative of f at a diagonal matrix is a Hadamard product with the Loewner matrix, and Chapter 13 §08 says that a Hadamard product of positive semidefinite matrices is positive semidefinite (§09). *Transfer:* to show that a matrix-valued function is increasing, differentiate it and recognize the derivative as an entrywise product.
- **A limit of Riemann sums carries the Loewner order through an integral.** The positive semidefinite cone is closed, so an integral of positive semidefinite matrices is positive semidefinite (§10). *Transfer:* an order defined by a closed convex cone survives any limit, so it survives integration; the only work is building the integral.
- **Lie’s product formula turns a sum in the exponent into a product.** [Lemma 21.12.1](#) replaces $e^{\mathbf{A}+\mathbf{B}}$, which nothing can be said about directly, by a limit of products, each of which can be attacked with Cauchy–Schwarz (this section). *Transfer:* when two operations fail to commute, split one of them into k small steps, estimate the error of a single step as $O(1/k^2)$, and let telescoping absorb the factor k .

The chapter paid the following promises, in the words in which they were made.

The promise	Made in	Paid by
Hardy–Littlewood–Pólya, “ $\mathbf{x} \prec \mathbf{y}$ exactly when $\mathbf{x} = \mathbf{D}\mathbf{y}$ for a doubly stochastic $\mathbf{D}$ ”	Chapter 19 §09	§01
“the general consequences of a majorization $\mathbf{x} \prec \mathbf{y}$, namely $\sum_i \phi(x_i) \leq \sum_i \phi(y_i)$ for every convex ϕ ”	Chapter 17 §07	§02 and §07

The promise	Made in	Paid by
“for the whole class of functions that reverse majorization” (the product is Schur-concave)	Chapter 17 §08	§02
“ $\mathbf{A} \mapsto \operatorname{tr} f(\mathbf{A}) \dots$ is proved in Chapter 21”	Chapter 18 §10	§02
“It has not said which norms are invariant under unitary multiplication, which is Chapter 21”	Chapter 16 §08	§04
“The repair is to apply N_k to the singular values instead of the eigenvalues, which is what Chapter 21 does”	Chapter 17 §06	§04
Eckart–Young “in every unitarily invariant norm”, by a route that “needs inequalities between the singular values of $\mathbf{A}$ and of $\mathbf{A} - \mathbf{B}$, not a projection argument”	Chapter 13 §11, Chapter 16 §08, Chapter 17 §09	§06
“ $\mathbf{W}$ also minimizes the distance in every unitarily invariant norm”	Chapter 13 §10	§06
“with $\phi(t) = t^2$ this is Corollary 20.5.4 again”	Chapter 20 §05	§07
“one statement covering every unitarily invariant norm” (the index-by-index singular-value bound and Mirsky)	Chapter 20 §10 and §11	§07
“ $t \mapsto t^s$ is operator monotone on $[0, \infty)$ exactly for $0 \leq s \leq 1$ ”, with “the companion fact that $t \mapsto \log t$ is operator monotone while $t \mapsto e^t$ is not”	Chapter 13 §05	§10, modulo (A7)

Some things the chapter did not do. Loewner’s theorem is proved here only up to one quoted fact, **(A7)**, the Nevanlinna–Pick characterization of Pick functions; §10 says precisely which implication rests on it, and §09’s elementary half — that for a continuously differentiable f , monotonicity of every order is exactly positivity of every Loewner matrix — is proved in full. The Lieb–Thirring inequality, which would remove the restriction to $m = 2^r$ in [Lemma 21.12.3](#), is not proved, and nothing here needs it. Lieb’s concavity theorem is stated above and not proved, and nothing in the book depends on it. The equality cases of von Neumann’s trace inequality and of Golden–Thompson were treated only in the directions that were needed.

Part V ends here. It began in Chapter 16 with a complaint: nine chapters had said “close” and “small” with nothing to measure by. A norm supplied the yardstick, and the five chapters since

have been spent finding out what a yardstick is good for. Chapter 17 made the eigenvalues of a Hermitian matrix into the values of an optimization problem, so that they could be compared. Chapter 18 supplied the language of convexity in which such comparisons live. Chapter 19 studied the one class of non-Hermitian matrices — the nonnegative ones — where a comparable theory exists, and produced Birkhoff’s theorem, which this chapter used once. Chapter 20 asked how much the answers move when the data move. And Chapter 21 has collected the inequalities that all five chapters kept needing, and shown that they are one inequality seen from different sides. What none of these chapters has done is compute anything: every proof here is an existence or a comparison, and not one of them is an algorithm. That is the subject of Chapter 24.

21.12.8 Exercises

21.12.8.1 A. Check your understanding

Exercise. A1

- State the Golden–Thompson inequality, with its hypotheses.
- State Lie’s product formula, and say for which matrices it holds.
- True or false: for Hermitian $\mathbf{A}, \mathbf{B}$ the matrix $e^{\mathbf{A}}e^{\mathbf{B}}$ is positive semidefinite. Justify your answer.
- True or false: $\operatorname{tr} e^{\mathbf{A}+\mathbf{B}} \leq \operatorname{tr}(e^{\mathbf{B}}e^{\mathbf{A}})$ for Hermitian $\mathbf{A}, \mathbf{B}$. Justify your answer.
- True or false: $\operatorname{tr} e^{\mathbf{A}+\mathbf{B}+\mathbf{C}} \leq \operatorname{tr}(e^{\mathbf{A}}e^{\mathbf{B}}e^{\mathbf{C}})$ for Hermitian $\mathbf{A}, \mathbf{B}, \mathbf{C}$. Justify your answer.
- Where in the proof of [Theorem 21.12.4](#) is the hypothesis that $\mathbf{A}$ and $\mathbf{B}$ are Hermitian used?

Solution.

- For Hermitian $\mathbf{A}, \mathbf{B} \in M_n(\mathbb{C})$, $\operatorname{tr} e^{\mathbf{A}+\mathbf{B}} \leq \operatorname{tr}(e^{\mathbf{A}}e^{\mathbf{B}})$ ([Theorem 21.12.4](#)). No commutativity is assumed; if $\mathbf{AB} = \mathbf{BA}$ the two sides are equal.
- $e^{\mathbf{A}+\mathbf{B}} = \lim_k (e^{\mathbf{A}/k}e^{\mathbf{B}/k})^k$ for **all** $\mathbf{A}, \mathbf{B} \in M_n(\mathbb{C})$, Hermitian or not ([Lemma 21.12.1](#)).
- False, since positive semidefinite includes Hermitian (Definition 13.1.1 (P1)) and $(e^{\mathbf{A}}e^{\mathbf{B}})^* = e^{\mathbf{B}}e^{\mathbf{A}}$, which equals $e^{\mathbf{A}}e^{\mathbf{B}}$ only when the two exponentials commute. In [Example \(The inequality on a two-by-two family\)](#) with $a = b = 1$,

$$e^{\mathbf{A}}e^{\mathbf{B}} = \begin{pmatrix} e \cosh 1 & e \sinh 1 \\ e^{-1} \sinh 1 & e^{-1} \cosh 1 \end{pmatrix},$$

which is not symmetric because $e \sinh 1 \neq e^{-1} \sinh 1$.

- True. $\operatorname{tr}(e^{\mathbf{B}}e^{\mathbf{A}}) = \operatorname{tr}(e^{\mathbf{A}}e^{\mathbf{B}})$ by Theorem 0.9.6 (3), and $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$, so this is [Theorem 21.12.4](#) with the roles exchanged.
- False; see [Example \(The three-matrix version fails\)](#), where the right side is negative and the left side positive. The right side need not even be real (the remark after that example).
- In two places. First, to know that $e^{\mathbf{A}/m}$ and $e^{\mathbf{B}/m}$ are positive semidefinite, which is the hypothesis of [Lemma 21.12.3](#) (b). Second, in [Lemma 21.12.2](#), which is what makes $\operatorname{tr}(e^{\mathbf{A}}e^{\mathbf{B}})$ real so that the inequality is a statement about real numbers.

21.12.8.2 B. Practice

Exercise. B1: Both sides, exactly

Let

$$\mathbf{A} = \begin{pmatrix} 0 & 0 \\ 0 & -2 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

- (a) Compute $e^{\mathbf{A}}$, $e^{\mathbf{B}}$, and $\text{tr}(e^{\mathbf{A}}e^{\mathbf{B}})$ exactly.
- (b) Compute $\text{tr}e^{\mathbf{A}+\mathbf{B}}$ exactly, and verify [Theorem 21.12.4](#) numerically to three decimal places.

Solution.

- (a) $\mathbf{A}$ is diagonal, so $e^{\mathbf{A}} = \text{diag}(1, e^{-2})$. Since $\mathbf{B}^2 = \mathbf{I}$, splitting the series into even and odd terms gives $e^{\mathbf{B}} = \cosh(1)\mathbf{I} + \sinh(1)\mathbf{B}$, that is

$$e^{\mathbf{B}} = \begin{pmatrix} \cosh 1 & \sinh 1 \\ \sinh 1 & \cosh 1 \end{pmatrix}.$$

Hence $e^{\mathbf{A}}e^{\mathbf{B}}$ has diagonal entries $\cosh 1$ and $e^{-2} \cosh 1$, so

$$\text{tr}(e^{\mathbf{A}}e^{\mathbf{B}}) = (1 + e^{-2}) \cosh 1 = \frac{(1 + e^{-2})(e + e^{-1})}{2}.$$

Numerically $(1 + 0.135335)(1.543081) = 1.751914$.

- (b) $\mathbf{A} + \mathbf{B} = \begin{pmatrix} 0 & 1 \\ 1 & -2 \end{pmatrix}$ has trace -2 and determinant -1 , so its characteristic polynomial is $x^2 + 2x - 1$ and its eigenvalues are $-1 \pm \sqrt{2}$. A Hermitian matrix has $\text{tr} e^{\mathbf{M}} = \sum_i e^{\lambda_i}$ by [Theorem 12.8.5 \(a\)](#) and [Theorem 9.3.3](#), so

$$\text{tr} e^{\mathbf{A}+\mathbf{B}} = e^{-1+\sqrt{2}} + e^{-1-\sqrt{2}} = 2e^{-1} \cosh \sqrt{2}.$$

Numerically $2(0.367879)(2.178183) = 1.602618$. Indeed $1.603 \leq 1.752$, and the inequality is strict because $\mathbf{A}$ and $\mathbf{B}$ do not commute: $\mathbf{AB} = \begin{pmatrix} 0 & 0 \\ -2 & 0 \end{pmatrix}$ while $\mathbf{BA} = \begin{pmatrix} 0 & -2 \\ 0 & 0 \end{pmatrix}$.

Exercise. B2: The trace power inequality on integers

Let $\mathbf{X} = \text{diag}(2, 1)$ and $\mathbf{Y} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$.

- (a) Check that $\mathbf{X}$ and $\mathbf{Y}$ are positive semidefinite, and compute $\text{tr}((\mathbf{XY})^m)$ and $\text{tr}(\mathbf{X}^m\mathbf{Y}^m)$ for $m = 1, 2, 4$.
- (b) Verify [Lemma 21.12.3 \(b\)](#) in these three cases, and say for which of them it is strict.

Solution.

- (a) $\mathbf{X}$ is diagonal, so its eigenvalues are 2 and 1; $\mathbf{Y}$ is symmetric with trace 2 and determinant 0, so its eigenvalues are 2 and 0. Both lists are nonnegative, so $\mathbf{X} \succeq \mathbf{O}$ and $\mathbf{Y} \succeq \mathbf{O}$ by [Theorem 13.1.4 \(b\)](#).

Now $\mathbf{XY} = \begin{pmatrix} 2 & 2 \\ 1 & 1 \end{pmatrix}$, which has trace 3 and determinant 0, so its eigenvalues are 3 and 0 and $\text{tr}((\mathbf{XY})^m) = 3^m$: the values are 3, 9, 81. On the other side $\mathbf{Y}^m = 2^{m-1}\mathbf{Y}$, since

$\mathbf{Y}^2 = 2\mathbf{Y}$, and $\mathbf{X}^m = \text{diag}(2^m, 1)$, so

$$\text{tr}(\mathbf{X}^m \mathbf{Y}^m) = 2^{m-1} \text{tr}(\text{diag}(2^m, 1) \mathbf{Y}) = 2^{m-1}(2^m + 1),$$

giving 3, 10, 136 for $m = 1, 2, 4$.

- (b) The inequality $3^m \leq 2^{m-1}(2^m + 1)$ holds in all three cases. It is an equality at $m = 1$, as the base case of the induction says it must be, and strict at $m = 2$ ($9 < 10$) and $m = 4$ ($81 < 136$).

Exercise. B3: Klein with the exponential

Let $\mathbf{A}, \mathbf{B} \in M_n(\mathbb{C})$ be Hermitian. Use [Theorem 21.12.6](#) with $f(t) = e^t$ to prove

$$\text{tr} e^{\mathbf{A}} \geq \text{tr} e^{\mathbf{B}} + \text{tr}((\mathbf{A} - \mathbf{B})e^{\mathbf{B}}),$$

and check the inequality for $\mathbf{A} = \text{diag}(1, 0)$, $\mathbf{B} = \mathbf{O}$.

Solution. The function $f(t) = e^t$ is differentiable on $I = \mathbb{R}$ with $f' = f$, and it is convex by Lemma 18.11.1 (a). Both spectra are real, hence contained in I . So [Theorem 21.12.6](#) gives

$$\text{tr}(e^{\mathbf{A}} - e^{\mathbf{B}} - (\mathbf{A} - \mathbf{B})e^{\mathbf{B}}) \geq 0,$$

and the trace is linear (Theorem 0.9.6 (1)), so this rearranges to the stated inequality.

For $\mathbf{A} = \text{diag}(1, 0)$ and $\mathbf{B} = \mathbf{O}$: $\text{tr} e^{\mathbf{A}} = e + 1 \approx 3.71828$, $\text{tr} e^{\mathbf{B}} = \text{tr} \mathbf{I} = 2$, and $\text{tr}(\mathbf{A}e^{\mathbf{O}}) = \text{tr} \mathbf{A} = 1$. So the inequality reads $3.71828 \geq 3$, and it is strict.

21.12.8.3 C. Going deeper

Exercise. C1: Two out of three is enough

Let $\mathbf{A}, \mathbf{B}, \mathbf{C} \in M_n(\mathbb{C})$ be Hermitian. Prove that

$$\text{tr} e^{\mathbf{A}+\mathbf{B}+\mathbf{C}} \leq \text{tr}(e^{\mathbf{A}}e^{\mathbf{B}}e^{\mathbf{C}})$$

holds whenever **some two** of the three matrices commute, treating the three cases separately. Explain why this does not contradict [Example \(The three-matrix version fails\)](#).

Hint: Theorem 10.9.4 (d), and the cyclic invariance of the trace.

Solution. Case 1: $\mathbf{BC} = \mathbf{CB}$. Then $e^{\mathbf{B}}e^{\mathbf{C}} = e^{\mathbf{B}+\mathbf{C}}$ by Theorem 10.9.4 (d), so $\text{tr}(e^{\mathbf{A}}e^{\mathbf{B}}e^{\mathbf{C}}) = \text{tr}(e^{\mathbf{A}}e^{\mathbf{B}+\mathbf{C}})$. Since $\mathbf{B} + \mathbf{C}$ is Hermitian, [Theorem 21.12.4](#) applied to the pair $\mathbf{A}, \mathbf{B} + \mathbf{C}$ gives $\text{tr} e^{\mathbf{A}+\mathbf{B}+\mathbf{C}} \leq \text{tr}(e^{\mathbf{A}}e^{\mathbf{B}+\mathbf{C}})$, as required.

Case 2: $\mathbf{AB} = \mathbf{BA}$. Then $e^{\mathbf{A}}e^{\mathbf{B}} = e^{\mathbf{A}+\mathbf{B}}$, so $\text{tr}(e^{\mathbf{A}}e^{\mathbf{B}}e^{\mathbf{C}}) = \text{tr}(e^{\mathbf{A}+\mathbf{B}}e^{\mathbf{C}})$, and [Theorem 21.12.4](#) applied to $\mathbf{A} + \mathbf{B}$ and $\mathbf{C}$ finishes.

Case 3: $\mathbf{AC} = \mathbf{CA}$. By Theorem 0.9.6 (3), $\text{tr}(e^{\mathbf{A}}e^{\mathbf{B}}e^{\mathbf{C}}) = \text{tr}(e^{\mathbf{C}}e^{\mathbf{A}}e^{\mathbf{B}}) = \text{tr}(e^{\mathbf{A}+\mathbf{C}}e^{\mathbf{B}})$, the last step because $\mathbf{A}$ and $\mathbf{C}$ commute. Now [Theorem 21.12.4](#) applied to $\mathbf{A} + \mathbf{C}$ and $\mathbf{B}$ gives

$$\operatorname{tr} e^{\mathbf{A}+\mathbf{B}+\mathbf{C}} \leq \operatorname{tr}(e^{\mathbf{A}+\mathbf{C}} e^{\mathbf{B}}).$$

There is no contradiction: in [Example \(The three-matrix version fails\)](#) no two of $\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3$ commute. Indeed, suppose two orthogonal projections $\mathbf{P}, \mathbf{Q}$ onto distinct lines commute. Then $\mathbf{PQ}$ is Hermitian, since $(\mathbf{PQ})^* = \mathbf{QP} = \mathbf{PQ}$, and idempotent, since $\mathbf{PQPQ} = \mathbf{P}^2\mathbf{Q}^2 = \mathbf{PQ}$; so it is an orthogonal projection, and its image lies in the image of $\mathbf{P}$ and in that of $\mathbf{Q}$, which meet only in $\mathbf{0}$. Hence $\mathbf{PQ} = \mathbf{0}$, which for $\mathbf{P} = \mathbf{u}\mathbf{u}^\top$ and $\mathbf{Q} = \mathbf{w}\mathbf{w}^\top$ says $\mathbf{u} \cdot \mathbf{w} = 0$. But the three unit vectors above pair to $-\frac{1}{2}$, never 0. For instance $\mathbf{P}_1\mathbf{P}_2 = \frac{1}{4}\begin{pmatrix} 1 & -\sqrt{3} \\ 0 & 0 \end{pmatrix} \neq \mathbf{0}$.

Exercise. C2: The symmetric Lie formula

Let $\mathbf{A}, \mathbf{B} \in M_n(\mathbb{C})$.

- Prove that $e^{\mathbf{A}+\mathbf{B}} = \lim_{k \rightarrow \infty} (e^{\mathbf{A}/2k} e^{\mathbf{B}/k} e^{\mathbf{A}/2k})^k$.
- Deduce that for Hermitian $\mathbf{A}, \mathbf{B}$ one also has $\operatorname{tr} e^{\mathbf{A}+\mathbf{B}} \leq \operatorname{tr}(e^{\mathbf{A}/2} e^{\mathbf{B}} e^{\mathbf{A}/2})$, and explain why this is not a new inequality.

Hint: for (a), rerun Steps 2–4 of the proof of [Lemma 21.12.1](#) with the new $\mathbf{T}_k$; only the constant M changes.

Solution.

- Keep $\mathbf{S}_k = e^{(\mathbf{A}+\mathbf{B})/k}$ and set $\mathbf{T}_k = e^{\mathbf{A}/2k} e^{\mathbf{B}/k} e^{\mathbf{A}/2k}$. Step 1 of the proof of [Lemma 21.12.1](#) is unchanged: $\mathbf{S}_k^k = e^{\mathbf{A}+\mathbf{B}}$. For Step 2, submultiplicativity and Theorem 16.6.1 (a) give $\|\mathbf{T}_k\|_2 \leq e^{a/2k} e^{b/k} e^{a/2k} = e^{c/k}$, the same bound as before, with $a = \|\mathbf{A}\|_2$, $b = \|\mathbf{B}\|_2$, $c = a + b$. For Step 3, write $e^{\mathbf{X}} = \mathbf{I} + \mathbf{X} + \mathbf{R}(\mathbf{X})$ as there and multiply out the three factors. Choosing the $\mathbf{I}$ from all three gives $\mathbf{I}$; choosing exactly one nonidentity term gives $\frac{\mathbf{A}}{2k} + \frac{\mathbf{B}}{k} + \frac{\mathbf{A}}{2k} = \frac{\mathbf{A}+\mathbf{B}}{k}$ together with the three remainders $\mathbf{R}(\frac{\mathbf{A}}{2k})$, $\mathbf{R}(\frac{\mathbf{B}}{k})$, $\mathbf{R}(\frac{\mathbf{A}}{2k})$. Every other term is a product of at least two nonidentity factors. Once $k \geq \max(1, c)$, each $\mathbf{R}$ has norm at most $c^2 e / (2k^2)$ by Theorem 16.6.1 (c), so every nonidentity factor has norm at most K/k with $K = c + c^2 e / 2$; a product of j of them therefore has norm at most K^j / k^j , which for $2 \leq j \leq 3$ and $k \geq 1$ is at most $\max(K^2, K^3) / k^2$. There are 27 terms in all. Subtracting from $\mathbf{S}_k = \mathbf{I} + \frac{\mathbf{A}+\mathbf{B}}{k} + \mathbf{R}(\frac{\mathbf{A}+\mathbf{B}}{k})$, the $\mathbf{I}$ and the linear terms cancel and what is left is a sum of at most 27 terms, each of norm at most a constant over k^2 , so $\|\mathbf{S}_k - \mathbf{T}_k\|_2 \leq M'/k^2$ for a constant M' depending only on $\mathbf{A}$ and $\mathbf{B}$. Step 4 is verbatim, and yields $\|e^{\mathbf{A}+\mathbf{B}} - \mathbf{T}_k^k\|_2 \leq e^c M' / k \rightarrow 0$.
- By Theorem 0.9.6 (3), $\operatorname{tr}(e^{\mathbf{A}/2} e^{\mathbf{B}} e^{\mathbf{A}/2}) = \operatorname{tr}(e^{\mathbf{A}/2} e^{\mathbf{A}/2} e^{\mathbf{B}}) = \operatorname{tr}(e^{\mathbf{A}} e^{\mathbf{B}})$, the last step by Theorem 10.9.4 (d) since $\mathbf{A}/2$ commutes with itself. So the two right-hand sides are the same number and the inequality is [Theorem 21.12.4](#) restated. What is new is only the form: the middle expression is a positive semidefinite matrix, being $(e^{\mathbf{A}/2})^* e^{\mathbf{B}} e^{\mathbf{A}/2}$ with $e^{\mathbf{A}/2}$ Hermitian (Proposition 13.1.2 (a)), whereas $e^{\mathbf{A}} e^{\mathbf{B}}$ is not Hermitian. The warning after [Example \(The inequality on a two-by-two family\)](#) shows that the Loewner-order strengthening of this form is nevertheless false.

Exercise. C3: When is the relative entropy zero?

Let $\mathbf{A}, \mathbf{B} \in M_n(\mathbb{C})$ be positive definite with $\operatorname{tr} \mathbf{A} = \operatorname{tr} \mathbf{B} = 1$.

- (a) Show that equality holds in [Lemma 21.12.5](#) for $f(t) = t \log t$ only when $s = t$.
 (b) Deduce that $\text{tr}(\mathbf{A}(\log \mathbf{A} - \log \mathbf{B})) = 0$ forces $\mathbf{A} = \mathbf{B}$.

Hint: for (b), look at which coefficients c_{ij} in the proof of [Theorem 21.12.6](#) can be nonzero.

Solution.

- (a) By [Lemma 21.12.7](#), $f(t) = t \log t$ is **strictly** convex on $(0, \infty)$. Let $s \neq t$ in $(0, \infty)$ and let $m = \frac{1}{2}(s + t)$ be their midpoint, which also lies in $(0, \infty)$. Strict convexity gives

$$f(m) < \frac{1}{2}f(t) + \frac{1}{2}f(s),$$

while [Lemma 21.12.5](#) applied at the point t with m in place of s gives

$$f(m) \geq f(t) + f'(t)(m - t) = f(t) + \frac{1}{2}f'(t)(s - t).$$

Combining and multiplying by 2, $2f(t) + f'(t)(s - t) < f(t) + f(s)$, that is $f(s) > f(t) + f'(t)(s - t)$. So the inequality is strict whenever $s \neq t$.

- (b) In the proof of [Theorem 21.12.6](#) with $f(t) = t \log t$, the total is $\sum_{i,j} c_{ij} [f(\lambda_i) - f(\mu_j) - f'(\mu_j)(\lambda_i - \mu_j)]$, with all $c_{ij} \geq 0$ and every bracket ≥ 0 . The computation in [Corollary 21.12.8](#) shows that this total equals $\text{tr}(\mathbf{A}(\log \mathbf{A} - \log \mathbf{B}))$ when $\text{tr} \mathbf{A} = \text{tr} \mathbf{B}$. If it is 0, then $c_{ij} = 0$ whenever the bracket is nonzero, that is, by (a), whenever $\lambda_i \neq \mu_j$.

So $\langle \mathbf{v}_j, \mathbf{u}_i \rangle = 0$ whenever $\lambda_i \neq \mu_j$. Fix j and expand $\mathbf{v}_j = \sum_i \langle \mathbf{v}_j, \mathbf{u}_i \rangle \mathbf{u}_i$: only indices i with $\lambda_i = \mu_j$ contribute, so $\mathbf{v}_j$ lies in the eigenspace of $\mathbf{A}$ for the eigenvalue μ_j , giving $\mathbf{A}\mathbf{v}_j = \mu_j \mathbf{v}_j = \mathbf{B}\mathbf{v}_j$. This holds for every j , and the $\mathbf{v}_j$ form a basis, so $\mathbf{A} = \mathbf{B}$.