1 Block magic

Recall that our goal is to prove that \((A, B) \mapsto S(A \| B)\) is jointly convex on positive matrices, where
\[
S(A \| B) = \text{Tr} \left( A(\log A - \log B) - (A - B) \right),
\]
and that we established the following result as a first step.

**Lemma 1.1.** The map \((A, B) \mapsto A : B\) is jointly concave on positive matrices, where
\[
A : B = (A^{-1} + B^{-1})^{-1} - 1.
\]

Our approach is via another concavity theorem of Lieb.

**Theorem 1.2.** For any matrix \(X \in \mathbb{M}_n(C)\) and \(0 \leq t \leq 1\), it holds that
\[
(A, B) \mapsto \text{Tr} \left( X^* A^t X B^{1-t} \right)
\]
is jointly concave on \(A, B > 0\).

Indeed, with **Theorem 1.2** in hand, let us define the function, for \(0 \leq t \leq 1\),
\[
I_t(A, X) := \text{Tr} \left( X^* A^t X A^{1-t} - X^* X A \right).
\]
Note that \(A \mapsto I_t(A, X)\) is concave by **Theorem 1.2** because \(A \mapsto \text{Tr}(X^* X A)\) is linear in \(A\).

Because \(I_0(A, X) = 0\), the function
\[
I(A, X) := \frac{d}{dt} \bigg|_{t=0^+} I_t(A, X) = \lim_{\varepsilon \to 0^+} \frac{I_\varepsilon(A, X) - I_0(A, X)}{\varepsilon} = \lim_{\varepsilon \to 0} \frac{I_\varepsilon(A, X)}{\varepsilon}
\]
is also concave in \(A\), since it is a limit of concave functions. Observe that
\[
\frac{d}{dt} A^t = \frac{d}{dt} e^{t \log A} = (\log A) A^t, \quad \frac{d}{dt} A^{1-t} = -(\log A) A^{1-t},
\]
hence
\[
I(A, X) = \text{Tr} \left( X^* (\log A) X A - X^* X (\log A) A \right)
\]
Now comes the amazing twist: Define the block matrices
\[
T = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}, \quad X = \begin{bmatrix} 0 & 0 \\ I & 0 \end{bmatrix}.
\]
Then \(I(T, X) = -S(A \| B)\). Since \(T \mapsto I(T, X)\) is concave, it follows that \((A, B) \mapsto S(A \| B)\) is jointly convex, establishing our goal. We are left to prove **Theorem 1.2**.

2 Tensor magic

We are left to prove **Theorem 1.2**. The magic is that it is equivalent to the assertion that the function
\[
F(A, B) = A^t \otimes B^{1-t}
\]
is jointly operator concave on \(A, B \geq 0\).
2.1 Tensor products of operators

The tensor product $\mathcal{H} = \mathbb{C}^m \otimes \mathbb{C}^n$ is an $mn$-dimensional complex inner product space spanned by the orthonormal basis $\{u_i \otimes v_j\}$, where $\{u_1, \ldots, u_m\}$ and $\{v_1, \ldots, v_n\}$ are any orthonormal bases for $\mathbb{C}^m$ and $\mathbb{C}^n$. For vectors $u, u' \in \mathbb{C}^m, v, v' \in \mathbb{C}^n$, the inner product is given by

$$\langle u \otimes v, u' \otimes v' \rangle_{\mathcal{H}} = \langle u, u' \rangle \langle v, v' \rangle.$$

Of course, the space $\mathcal{H}$ contains more vectors than just those of the form $u \otimes v$. For instance, $u_1 \otimes u_2 + u_2 \otimes u_1$. The inner product is extended to such vectors by linearity. As an inner product space, $\mathcal{H}$ is isomorphic to $\mathbb{C}^{mn}$.

We can represent this isomorphism as follows: Given $u \in \mathbb{C}^m, v \in \mathbb{C}^n$, there is a canonical representation of $u \otimes v$ as an $m \times n$ matrix:

$$(u \otimes v)_{ij} = u_i v_j.$$

Moreover, it holds that $\langle u \otimes v, u' \otimes v' \rangle_{\mathcal{H}} = \text{Tr} ((u \otimes v)'(u' \otimes v'))$, where in the second expression we consider the 2-tensors as matrices.

Note that the space $\mathbb{C}^m$ is isomorphic to the collection of univariate polynomials of degree at most $m - 1$: $p(x) = a_{m-1}x^{m-1} + a_{m-2}x^{m-2} + \cdots + a_1 x + a_0$, where $x$ is a complex variable. It holds that $\mathbb{C}^m \otimes \mathbb{C}^n$ is isomorphic to the collection of bivariate polynomials of degree at most $m - 1$ in $x$ and $n - 1$ in $y$: $q(x, y) = \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} a_{ij} x^i y^j$.

If we have linear operators $A : \mathbb{C}^m \to \mathbb{C}^{m'}$ and $B : \mathbb{C}^n \to \mathbb{C}^{n'}$, one defines the linear operator $A \otimes B : \mathbb{C}^m \otimes \mathbb{C}^n \to \mathbb{C}^{m'} \otimes \mathbb{C}^{n'}$ by

$$(A \otimes B)(x \otimes y) = (Ax) \otimes (By),$$

with the extension of to all of $\mathbb{C}^m \otimes \mathbb{C}^n$ given by linearity.

If $A$ and $B$ are represented by $m \times m'$ and $n \times n'$ matrices, respectively, then we can represent $A \otimes B$ as the $mn \times m'n'$ matrix given by

$$(A \otimes B)_{ij, i'j'} = A_{ii'}B_{jj'}.$$

If additionally we have $m = n$ and $m' = n'$ (so that both $A$ and $B$ are square matrices), then

$$\text{Tr}(A \otimes B) = \sum_{i=1}^{m} \sum_{j=1}^{n} (A \otimes B)_{ij, ij} = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ii}B_{jj} = \text{Tr}(A)\text{Tr}(B).$$

If we recall the Schur product $A \circ B$ given by $(A \circ B)_{ij} = A_{ij}B_{ij}$, then we can confirm our earlier claim that $A \circ B$ is a principal submatrix of $A \otimes B$:

$$A \circ B = (A \otimes B)_{ii, jj}.$$

2.2 Ando’s identity

Lemma 2.1. Consider $A \in \mathbb{H}_m$ and $B \in \mathbb{H}_n$, and $X \in \mathbb{M}_{m \times n}(\mathbb{C})$ arbitrary. Then,

$$\langle X, (A \otimes B)X \rangle = \text{Tr}(X^*AXB),$$

where on the LHS, $X$ is considered as a vector in $\mathbb{C}^m \otimes \mathbb{C}^n$, and $\bar{B}$ denotes the matrix whose entries are the complex conjugates of those in $B$. 

2
Proof. Considered as a matrix, we have

\[(A \otimes B)X\]_{ij} = \sum_{k,\ell} A_{ik} X_{k\ell} B_{j\ell}.

Since $B$ is Hermitian, we can write the latter sum as

\[\sum_{k,\ell} A_{ik} X_{k\ell} \bar{B}_{\ell j} = [AX\bar{B}]_{ij}.

Hence:

\[\langle X, (A \otimes B)X \rangle = \text{Tr}(X^\ast (A \otimes B)X) = \text{Tr}(X^\ast AX\bar{B}). \quad \square\]

We conclude that

\[\text{Tr}(X^\ast AX\bar{B}) \geq 0 \quad \forall X \in \mathbb{M}_{m \times n}(\mathbb{C}) \iff A \otimes B \succeq 0.\]

In particular, we have the following corollary.

**Corollary 2.2.** The mapping $(A, B) \mapsto \text{Tr}(X^\ast AX\bar{B})$ is jointly concave on positive matrices for every $X \in \mathbb{M}_{m \times n}(\mathbb{C})$ if and only if the mapping $(A, B) \mapsto A^p \otimes B^q$ is jointly operator concave on PSD matrices.

Thus we reduce Theorem 1.2 to the following theorem.

**Theorem 2.3.** For every $0 \leq t \leq 1$, the mapping $(A, B) \mapsto A^t \otimes B^{1-t}$ is jointly operator concave on PSD matrices.

Proof. Recall the representation proved in Lecture 4: For $0 < p < 1$, we have

\[A^p = C_p \int_0^\infty t^p A(t + A)^{-1} dt = C_p \int_0^\infty t^p ((A/t)^{-1} + I)^{-1} dt = C_p \int_0^\infty t^p \left( A/t : I \right) dt.

Note that $A^t \otimes B^{1-t} = (I \otimes B)(A \otimes B^{-1})^t$, so using this representation, we can write

\[A^t \otimes B^{1-t} = C_p \int_0^\infty t^p (I \otimes B) \left( (A \otimes B^{-1})/t : I \right) dt = C_p \int_0^\infty t^p \left( (A \otimes I)/t : I \otimes B \right) dt.

Thus we are done by Lemma 1.1 since the mapping $(A, B) \mapsto A : B$ is jointly operator concave on positive matrices. \[ \square \]

3 Discussion

3.1 Extension of the Golden-Thompson inequality to three matrices

Given a function $F$, let us define the Fréchet derivative (when it exists) by

\[D_K F(A) = \frac{d}{dt} \bigg|_{t=0} F(A + tK).

For $A > 0$ and $K$ Hermitian, define the operator

\[T_A(K) := D_K \log(A).\]
Note that
\[
\log(x) = \int_1^x \frac{1}{t} \, dt = \int_0^{x-1} \frac{1}{1+t} \, dt = \int_0^\infty \frac{1}{1+t} - \frac{1}{x+t} \, dt,
\]
hence for a Hermitian matrix \( A \),
\[
\log(A) = \int_0^\infty (1 + t)^{-1}I - (tI + A)^{-1} \, dt.
\]
Recalling that \( D_K A^{-1} = -A^{-1}KA^{-1} \), we calculate
\[
T_A(K) = D_K \log(A) = \int_0^\infty (tI + A)^{-1}K(tI + A)^{-1} \, dt.
\]
In particular, this implies that \( T_A \) is a positive operator: \( T_A(K) \geq 0 \) when \( K \geq 0 \). (This is equivalent to the fact we established earlier that \( A \mapsto \log(A) \) is operator monotone.)

**Theorem 3.1** (Lieb). For all \( A, B, C \in H^n \), it holds that
\[
\text{Tr}(e^{A+B+C}) \leq \text{Tr}\left(e^C \text{Tr}_{e^{-A}(e^B)}\right).
\]
This simplifies to \( \text{Tr}(e^{A+B+C}) \leq \text{Tr}(e^A e^B e^C) \) if \( A \) and \( B \) commute.

Note that if \( A, B \) commute, then \( \text{Tr}_{e^{-A}}(e^B) = e^{A+B} \) since \( \frac{d}{dt} \bigg|_{t=0} \log(e^{-a} + te^b) = e^{a+b} \) holds for numbers. In particular, taking \( B = 0 \) recovers the Golden-Thompson inequality \( \text{Tr}(e^{A+C}) \leq \text{Tr}(e^A e^C) \). We will need two basic results.

**Lemma 3.2.** If \( f : I \rightarrow \mathbb{R} \) is continuously differentiable on an open interval \( I \) and \( \text{spec}(A) \subseteq I \), then
\[
\frac{d}{dt} \bigg|_{t=0} \text{Tr}(f(A + tB)) = \text{Tr}(f'(A)B).
\] (3.1)

This follows from “first order perturbation theory” which is the following simple idea: Consider the case where \( f(A) = A^n \), \( n \geq 1 \) an integer. Then as \( \varepsilon \rightarrow 0 \),
\[
\text{Tr}((A + \varepsilon B)^n - A^n) = \varepsilon \text{Tr}\left(BA^{n-1} + ABA^{n-2} + \cdots + A^nB\right) + O(\varepsilon^2) = \varepsilon n \text{Tr}(BA^{n-1}) + O(\varepsilon^2),
\]
hence (3.1) holds. More generally, this shows that (3.1) holds when \( f \) is any polynomial. Now approximating \( f \) by polynomials yields the general claim.

**Lemma 3.3.** Let \( 'C \) be a convex cone, and consider a differentiable convex function \( F : 'C \rightarrow \mathbb{R} \) that is 1-homogeneous in the sense that \( F(\lambda A) = \lambda F(A) \) for \( \lambda > 0 \). Then,
\[
D_B(A) \leq F(B), \quad \forall A, B \in 'C.
\]

**Proof.** Since \( F \) is 1-homogeneous and convex, it holds that
\[
F(A + B) = 2 \cdot F((A + B)/2) \leq 2 \cdot (F(A)/2 + F(B)/2) = F(A) + F(B),
\]
hence \( F(A + tB) - F(A) \leq F(tB) = tF(B) \), and the result follows by taking \( t \rightarrow 0 \). \( \square \)
Recall that Lieb’s concavity theorem asserts that for any Hermitian \( H \), the map \( X \mapsto \operatorname{Tr}(e^{H+X}) \) is concave on positive matrices, hence the map \( X \mapsto -\operatorname{Tr}(e^{H+X}) \) is convex and 1-homogeneous. Therefore Lemma 3.3 gives

\[
\operatorname{Tr} \left( e^{H+X} \right) \leq \frac{d}{dt} \bigg|_{t=0} \operatorname{Tr} \left( e^{H+X+tY} \right) = \left( e^{H+X} Y e^{H+X} \right) = \operatorname{Tr} \left( e^{H+X} T_X(Y) \right),
\]

for \( H \in \mathbf{H}_n \) and \( X, Y > 0 \).

Thus taking \( H = A + C, X = e^{-A} \) and \( Y = e^B \) gives

\[
\operatorname{Tr}(e^{A+B+C}) \leq \operatorname{Tr} \left( e^{C} T_{e^{-A}}(e^B) \right),
\]

proving Theorem 3.1.

### 3.2 Integral representations

In the proof of Theorem 2.3, we once again used the integral representation: For \( 0 < p < 1 \) and \( a > 0 \),

\[
a^p = C_p \int_0^\infty t^p \frac{a}{t+a} dt,
\]

where \( C_p \) is a constant depending on \( p \).

In fact, a generalization of this approach is the only way to establish operator concavity. Let us state a characterization of Loewner. Define \( \mathcal{M}_\infty(0, \infty) \) as the set of functions \( f : (0, \infty) \to \mathbb{R} \) that are matrix monotone, i.e., satisfy \( A \succeq B \implies f(A) \succeq f(B) \) for \( \text{spec}(A), \text{spec}(B) \subseteq (0, \infty) \).

Denote the upper complex halfplane by \( \mathbb{C}_+ := \{ z \in \mathbb{C} : \text{Im} z > 0 \} \). Let \( \mathcal{P} \) denote the class of Pick functions: Analytic functions \( f : \mathbb{C}_+ \to \mathbb{C}_+ \). Examples include \( f(z) = z^p \) for \( p \leq 1 \), noting that if \( z = re^{i\theta} \), then \( z^p = re^{ip\theta} \). This class also includes the principal branch of the logarithm \( f(z) = \log r + i\theta \).

**Theorem 3.4 (Loewner).** The following are equivalent:

(a) \( f \in \mathcal{M}_\infty(0, \infty) \)

(b) There exists a positive measure \( \mu \) on \( (0, \infty) \) and \( \beta \geq 0 \) such that

\[
f(x) = f(0) + \beta x + \int_0^\infty \frac{t x}{x + t} \ d\mu(t).
\]

(c) \( f \) admits an analytic extension to a Pick function on \( \mathbb{C}_+ \cup (0, \infty) \).

(d) \( f \) is matrix concave on \( (0, \infty) \).

Observe that (b) \( \implies (a) \land (d) \): Note that \( \frac{x}{x+t} = 1 - \frac{t}{x+t} \), hence for Hermitian \( A \), (b) gives

\[
f(A) = f(0) + \beta A + \int_{-1}^1 t \left( 1 - t(A + tI)^{-1} \right) \ d\mu(t).
\]

Since the map \( A \mapsto -(A + tI)^{-1} \) is matrix monotone, it follows that \( f \in \mathcal{M}_\infty(0, \infty) \). Since the map \( A \mapsto (A + tI)^{-1} \) is matrix convex, it follows that \( f \) is matrix concave. The hard direction of the theorem is (a) \( \implies (c) \), and we will not go into it here (see B. Simons’ extensive book _Loewner’s Theorem on Monotone Matrix Functions_).
**Dirichlet problem on the unit disk.** But we can still understand why complex analysis arises naturally. Let \( \mathbb{D} := \{ z \in \mathbb{C} : |z| < 1 \} \) denote the open unit disk, and suppose that \( g : \partial \mathbb{D} \to \mathbb{R} \) is a continuous function on the boundary. Then there is a unique harmonic function \( u : \overline{\mathbb{D}} \to \mathbb{R} \) such that \( u|_{\partial \mathbb{D}} = g \). Harmonic means that \( u(z) \) is given by its averages on small circles around \( z \): As \( \varepsilon \to 0 \),

\[
 u(z) = \frac{1}{2\pi} \int u(z + \varepsilon e^{i\theta}) \, d\theta + O(\varepsilon).
\]

If we let \( \{B_t\} \) denote a 2-dimensional Brownian motion, and let \( T \) denote the first time at which \( B_t \) exits \( \mathbb{D} \), then for \( z \in \mathbb{D} \),

\[
 u(z) = \mathbb{E}[g(B_T) \mid B_0 = z].
\]

In other words, \( u \) is specified by the boundary values, and its value at \( z \) is prescribed by the distribution of where Brownian motion started at \( z \) exits the disk. Fortunately, this has a nice expression given by the Poisson integral formula.

Since \( u \) is harmonic, it holds that

\[
 u(0) = \frac{1}{2\pi} \int g(e^{i\theta}) \, d\theta.
\]

In other words, the exit measure for \( B_t \) started at the origin is uniform on the unit circle, as we would expect. Define the linear fractional transformation

\[
 T_z(w) := \frac{w + z}{1 + \bar{z}w}.
\]

This is a conformal mapping that sends \( \overline{\mathbb{D}} \) to \( \overline{\mathbb{D}} \) and \( 0 \mapsto z \). If we define \( v_z(w) := u(T_z(w)) \), then \( v_z \) is also harmonic. (Intuitively, this is because a conformal map sends infinitesimal circles around 0 to infinitesimal circles around \( z \).) Therefore for \( z \in \mathbb{D} \), we can write

\[
 u(z) = v_z(0) = \frac{1}{2\pi} \int g(T_z(e^{i\theta})) \, d\theta.
\]

If we denote \( T_z(e^{i\theta}) = e^{i\phi} \), then \( e^{i\phi} = T_{-z}(e^{i\theta}) \), and substituting in the integral gives

\[
 u(z) = \frac{1}{2\pi} \int g(e^{i\phi}) \frac{d\theta}{d\phi} \, d\phi = \int \frac{1}{2\pi} \int \frac{1 - |z|^2}{|1 - z e^{-i\phi}|^2} \, g(e^{i\phi}) \frac{d\phi}{2\pi}.
\]

The kernel describes the exit measure of 2-dimensional Brownian motion (started at \( z \)) from the unit disk. (The pedagogical value of introducing the Brownian motion is that the “physical” formulation seems more intuitive.) Note that if we additionally assume that \( g \) is nonnegative, then we obtain \( u(z) \) as an integral of the Poisson kernel against a positive measure.

**The analytic perspective.** In the plane, every harmonic function is the real part of an analytic function. Suppose that a function \( f \) is analytic on an open neighborhood \( \mathcal{D} \) of \( \overline{\mathbb{D}} \). This means that \( f \) can be written as a convergent power series about every point in \( \mathcal{D} \). Or, equivalently, that the complex derivative \( \frac{df}{dz} \) exists for all \( z \in \mathcal{D} \).

In this case, \( f \) satisfies the Cauchy-Riemann equations which, in particular, imply that \( \text{Re} \, f \) and \( \text{Im} \, f \) are harmonic functions. We can use the preceding discussion with \( u = \text{Re} \, f \) to write

\[
 \text{Re} \, f(z) = \int \frac{1 - |z|^2}{|1 - z e^{-i\phi}|^2} \, \text{Re} \, f(e^{i\phi}) \frac{d\phi}{2\pi} = \int \text{Re} \, K(e^{i\phi}, z) \, \text{Re} \, f(e^{i\phi}) \frac{d\phi}{2\pi}.
\]

6
where we define
\[ K(e^{i\theta}, z) := \frac{1 + ze^{-i\theta}}{1 - ze^{-i\theta}}. \]

Note that
\[ \Re\left( f(z) - \int K(e^{i\phi}, z) \Re f(e^{i\phi}) \frac{d\phi}{2\pi} \right) = 0. \]

Since the quantity in parentheses is an analytic function whose real part vanishes, the Cauchy-Riemann equations dictate that this quantity is constant on \( \overline{\mathbb{D}} \), hence we obtain a representation for \( f \) on \( \overline{\mathbb{D}} \):
\[ f(z) = i \Im f(0) + \int K(e^{i\phi}, z) \Re f(e^{i\phi}) \frac{d\phi}{2\pi}. \]

If \( \Re f > 0 \), we have
\[ f(z) = i \Im f(0) + \int K(e^{i\phi}, z) d\nu(\phi), \]
for some positive measure \( d\nu \) on \( \partial \mathbb{D} \). (This is called the Herglotz representation on \( \mathbb{D} \).)

Recall that we’re concerned with functions \( g : (0, \infty) \rightarrow \mathbb{R} \). Let’s suppose that \( g \) is the restriction of a function \( f : \mathbb{C} \rightarrow \mathbb{C} \) that is analytic on \( \mathbb{C}_+ \cup (0, \infty) \). If \( T : \mathbb{C}_+ \cup \{\infty\} \rightarrow \mathbb{D} \) is a conformal map, then we can obtain integral representations of \( f \) by applying the preceding argument to \( f \circ T^{-1} \), with the caveat that we need \( \Re(f \circ T^{-1}) > 0 \). This is where the class of Pick functions \( \mathcal{P} \) come into play: If \( f : \mathbb{C}_+ \rightarrow \mathbb{C}_+ \), then we have \(-if \circ T^{-1} : \mathbb{D} \rightarrow \{ z \in \mathbb{C} : \Re z > 0 \} \).

Applying this to \(-if \circ T^{-1} \) and working through the computation gives
\[ f(z) = \Re f(i) + i \int K(T(t), T(z)) d\mu(t), \]
where \( \mu \) is the pullback of the measure \( \nu \) under \( T \), i.e., \( \mu(A) = \nu(T[A]) \).

If \( t \neq \infty \), then
\[ i K(T(t), T(z)) = \frac{1 + tz}{t - z}. \]

And since \( T(\infty) = 1 \), it holds that \( i K(T(\infty), T(z)) = z \). It follows that
\[ f(z) = \Re f(i) + \mu(\mathbb{R})z + \int \frac{1 + tz}{t - z} d\mu(t). \]

While this is not quite the form of Theorem 3.4(c), one can get there with some further manipulation. (Briefly: It turns out that \( \mu((0, \infty)) = 0 \), hence the integral can be restricted to \((-\infty, 0)\).)