Matrix exponential, Lieb's inequality, proof of connector lemma

Lecturer : Aaditya Ramdas

1 Spectral decomposition of Hermitian matrices $\mathcal{H}_d$

They are a generalization of real-symmetric matrices to complex values: they satisfy the property that $A^* = A$, where A^* is the conjugate-transpose of the matrix A. For the standard Euclidean inner-product, this implies that $\langle Ax, y\rangle = \langle x, Ay\rangle$.

As a result of the spectral theorem, Hermitian matrices can be diagonalized, and the eigenvalues are all real. Let $V_\lambda := \{v : Av = \lambda v\}$ be the subspace formed by (linearly independent) eigenvectors with eigenvalue λ, and let P_λ denote the orthogonal projection onto this subspace. Then, one can write A in terms of its spectral decomposition:

$$A = \sum_{i=1}^{d} \lambda_i P_{\lambda_i}.$$

2 Functions on matrices

The above spectral theorem is useful because one can extend functions over the reals to functions of Hermitian matrices as

$$f(A) = \sum_i f(\lambda_i) P_{\lambda_i}.$$

For any interval $I \subseteq \mathbb{R}$, a function $f : I \mapsto \mathbb{R}$ is operator monotone if $A \preceq B$ implies that $f(A) \preceq f(B)$, it is operator convex if $f(\lambda A + (1-\lambda)B) \preceq \lambda f(A) + (1-\lambda)f(B)$, and it is operator concave if $-f$ is operator convex).

Theorem 1 (Lowner-Heinz) *(a) For $-1 \le p \le 0$, the function $f(t) = -t^p$ is operator monotone and operator concave.*

(b) For $0 \le p \le 1$, the function $f(t) = t^p$ is operator monotone and operator concave.

(c) For $1 \le p \le 2$, the function $f(t) = t^p$ is operator convex.

(d) The function $f(t) = \log t$ is operator monotone and operator concave, while $f(t) = t \log t$ is operator convex.

Given any function f, the corresponding trace function is given by $A \mapsto Tr(f(A)) = \sum_j f(\lambda_j)$. The trace function preserves monotonicity and convexity:

Lemma 2 (Trace-function preservation lemma) *Let $f : \mathbb{R} \mapsto \mathbb{R}$ be continuous. If the function $t \mapsto f(t)$ is monotone/convex/strictly-convex, then $A \mapsto Tr(f(A))$ is also monotone/convex/strictly-convex on $\mathcal{H}_d$.*

3 Matrix exponential

Unless otherwise mentioned, all matrices will be $d \times d$. Recall that $A \preceq B$ denotes the positive semidefinite (psd) ordering, and $\mathcal{S}_+$ denotes the psd cone. Define the exponential of a matrix as

$$\exp(A) := \sum_k \frac{A^k}{k!}.$$

The exponential of a matrix effectively exponentiates its eigenvalues; that is, if $A = PDP^{-1}$ is the eigenvalue decomposition of A, then we have:

$$\exp(A) = P\mathrm{diag}(\exp(d_i))P^{-1} = P\exp(D)P^{-1}.$$

Some properties include: $\exp(0) = I$, $\exp(A)^T = \exp(A^T)$, and $\exp(A^*) = \exp(A)^*$ (where A^* is the conjugate transpose of A). If X and Y commute, that is $XY = YX$, then $\exp(X)\exp(Y) = \exp(X + Y)$. Hence $\exp(X)\exp(-X) = I$, and so the matrix exponential is always invertible. Jacobi's formula implies that

$$\det(\exp(A)) = \exp(Tr(A)).$$

The matrix exponential appears naturally in the solution of ODEs. Indeed, the solution to $\dot{y}(t) = Ay(t), y(0) = 0$ is given by $y(t) = \exp(At)y_0$.

The matrix exponential results in a psd matrix. While the trace-exponential is monotone and strictly convex, the matrix exponential is neither operator monotone nor operator convex.

4 Lieb and Golden-Thompson

Theorem 3 (Lieb) *For any fixed Hermitian matrix H, the function $A \mapsto Tr\exp(H+\log A)$ is concave on $\mathcal{S}_+$.*

Theorem 4 (Golden-Thompson) *For Hermitian matrices A, B, we have*

$$Tr(\exp(A + B)) \leq Tr(\exp(A)\exp(B)).$$

4.1 Proof of Connector Lemma

Suppose (Y_t) is sub-ψ with self-normalizing process (U_t) and variance process (W_t). Fixing $\lambda \in [0, \lambda_{\max})$, Lieb's theorem and Jensen's inequality together imply

$$\mathbb{E}_{t-1}\mathrm{Tr}\exp\{\lambda Y_t - \psi(\lambda)\cdot(U_t + W_t)\} \leq \mathrm{Tr}\exp\{\lambda Y_{t-1} - \psi(\lambda)\cdot(U_{t-1} + W_t) + \log \mathbb{E}_{t-1} e^{\lambda \Delta Y_t - \psi(\lambda)\cdot \Delta U_t}\}.$$

Now we apply the sub-ψ property to the expectation, using the monotonicity of the matrix logarithm and trace exponential to obtain

$$\mathbb{E}_{t-1}\mathrm{Tr}\exp\{\lambda Y_t - \psi(\lambda)\cdot(U_t + W_t)\} \leq \mathrm{Tr}\exp\{\lambda Y_{t-1} - \psi(\lambda)\cdot(U_{t-1} + W_{t-1})\}.$$

This shows that the process $L_t := \mathrm{Tr}\exp\{\lambda Y_t - \psi(\lambda)\cdot(U_t + W_t)\}$ is a supermartingale, with $L_0 = d$. Next we show that $L_t \geq \exp\{\lambda\gamma_{\max}(Y_t) - \psi(\lambda)\gamma_{\max}(U_t + W_t)\}$ a.s. for all t, which is the canonical assumption. We repeat a short argument from Tropp (2012). First, by the monotonicity of the trace exponential,

$$\begin{aligned}\mathrm{Tr}\exp\{\lambda Y_t - \psi(\lambda)\cdot(U_t + W_t)\} &\geq \mathrm{Tr}\exp\{\lambda Y_t - \psi(\lambda)\gamma_{\max}(U_t + W_t)I_d\} \\ &\geq \gamma_{\max}(\exp\{\lambda Y_t - \psi(\lambda)\gamma_{\max}(U_t + W_t)I_d\}) =: B.\end{aligned}$$

using the fact that the trace of a positive semidefinite matrix is at least as large as its maximum eigenvalue. Then the spectral mapping property gives

$$B = \exp\{\gamma_{\max}(\lambda Y_t - \psi(\lambda)\gamma_{\max}(U_t + W_t)I_d).$$

Finally, we use the fact that $\gamma_{\max}(A - cI_d) = \gamma_{\max}(A) - c$ for any $A \in \mathcal{H}^d$ and $c \in \mathbb{R}$ to see that $B = \exp\{\lambda\gamma_{\max}(Y_t) - \psi(\lambda)\gamma_{\max}(U_t + W_t)\}$, completing the argument.

References

Tropp, J. A. (2012), 'User-friendly tail bounds for sums of random matrices', *Foundations of Computational Mathematics* **12**(4), 389–434.