An Introduction on Non-Commutative Functional Analysis: Quantised Calculus. Lecture 1: Matrix norms

AMSI Summer School, 2018.

We work exclusively over the field of complex numbers $\mathbb{C}$. Suppose that $d \geq 1$ and let $M_d = M_d(\mathbb{C})$ denote the $\mathbb{C}$-vector space of $d \times d$ matrices. Let 1 denote the $d \times d$ identity matrix, or 1_d if disambiguation is necessary.

Matrices as quantised numbers

There is a "quantum mechanical" perspective on matrix analysis which we will adopt in this course. We can view a $d \times d$ matrix A as being a "fuzzy number".

The matrix A has a "spectrum of possible values", defined to be the eigenvalues $\{\lambda_0(A), \lambda_1(A), \ldots, \lambda_{d-1}(A)\}$. In $d = 1$ these "fuzzy numbers" become "sharp" and take just one value.

According to this perspective:

- Adjoint of a matrix corresponds to complex conjugate of a number
- Self-adjoint matrix corresponds to a real number

The fact that matrix multiplication does not commute implies a kind of "uncertainty principle" for matrices, but we will not explore that in this lecture.

A motivating problem for matrix analysis

If $x \in \mathbb{C}$ and $|x| < 1$, we know that:

$$(1 - x)^{-1} = 1 + x + x^2 + \cdots$$

If $A \in M_d$, this suggests that $(1 - A)^{-1}$ may be computed as:

$$(1 - A)^{-1} = 1 + A + A^2 + \cdots$$

when is such a formula valid?

Another motivating problem for matrix analysis

Another example is the following: let $A \in M_d$ and $x : [0, T] \to \mathbb{C}^d$. Solve the initial value problem:

$$x'(t) = Ax(t), \quad t \in [0, T]$$

with $x(0) \in \mathbb{C}^d$ fixed. How do we solve this equation when $d > 1$? The solution (as in the scalar case) can be given as a convergent series:

$$x(t) = \exp(At)x(0) = \sum_{k \geq 0} \frac{A^k x(0)}{k!} t^k$$

How can we prove that this series actually converges?
In this lecture we will introduce the tools allowing for meaningful answers to these questions.

Notation

Let $\langle \cdot, \cdot \rangle$ denote the standard inner product of $\mathbb{C}^d$, conjugate linear in the first argument:

$$\langle x, y \rangle = \sum_{j=0}^{d-1} \overline{x_j} y_j$$

where $x = (x_0, \ldots, x_{d-1}), y = (y_0, \ldots, y_{d-1}) \in \mathbb{C}^d$. The notation e_j, $j = 0, \ldots, d-1$ denotes the jth standard basis vector.
The Euclidean norm is defined as:

$$\|x\|_2 := \sqrt{\langle x, x \rangle}.$$

Here, and from now on, the notation $\begin{pmatrix} \lambda_0 & & \\ & \ddots & \\ & & \lambda_{d-1} \end{pmatrix}$ means that the remaining entries are zeros.

The spectral theorem

A very powerful tool in matrix theory is the spectral theorem.
A matrix $A \in M_d$ is called self-adjoint if $A = A^*$ (recall that A^* is the conjugate transpose of A). So self-adjointness means that $\langle e_j, Ae_k \rangle = \langle Ae_j, e_k \rangle$. Recall that a matrix U is called unitary if $U^{-1} = U^*$.

Theorem 1.1 (The spectral theorem)

A self-adjoint matrix A can be unitarily diagonalised. That is, there exists a unitary matrix $U \in M_d$ such that

$$A = U \begin{pmatrix} \lambda_0 & & \\ & \ddots & \\ & & \lambda_{d-1} \end{pmatrix} U^*.$$

where $\{\lambda_0, \ldots, \lambda_{d-1}\}$ are the eigenvalues of A.

The Löwner partial order

Note that if A is self-adjoint, and $x \in \mathbb{C}^d$ then $\langle x, Ax \rangle$ is real.
A matrix $A \in M_d(\mathbb{C})$ is called positive semidefinite if:

$$\langle x, Ax \rangle \geq 0$$

for all $x \in \mathbb{C}^d$.
Given two self-adjoint matrices A and B, we say that $A \geq B$ if $A - B$ is positive semidefinite. This is called the Löwner partial ordering.
This is the quantum analogue of the usual ordering of real numbers, and a positive semidefinite matrix is like a positive number.

Properties of the Löwner partial order

Exercise 1.1

Show that a self-adjoint matrix A is positive semidefinite if and only if all its eigenvalues are non-negative.

Exercise 1.2

A positive semidefinite matrix is self-adjoint. (Hint: let $z = x + iy$ where $x, y \in \mathbb{C}^d$ and consider (z, Az).)

Exercise 1.3

Show that $A^*A \geq 0$ for any matrix A.

The Löwner partial order

The Löwner ordering of matrices can be counterintuitive at first sight.

Exercise 1.4

All of the following assertions are false in general:

(a) *If* $A \geq B$ *then* $A^2 \geq B^2$,

(b) *Either* $A \geq B$ *or* $B \geq A$,

(c) *If* $A \geq 0$ *and* $B \geq 0$, *then* $AB \geq 0$.

(d) *If* $A, B \geq 0$ *then* $2AB \leq A^2 + B^2$.

Find counterexamples for each of them (Hint: start by checking with 2×2 *matrices).*

Square root of a matrix

A square root B of a positive-semidefinite matrix A is a positive-semidefinite matrix B such that $B^2 = A$.

Theorem 1.2

Every positive-semidefinite matrix A has a unique square root.

To show that a square root exists, use the spectral theorem:

$$A = U \Lambda U^*$$

where

$$\Lambda = \begin{pmatrix} \lambda_0 & & \\ & \ddots & \\ & & \lambda_{d-1} \end{pmatrix}$$

is the diagonal matrix of eigenvalues.

Square root of a matrix

Since the eigenvalues of a positive-semidefinite matrix are non-negative, we can define:

$$B := U \begin{pmatrix} \lambda_0^{1/2} & & \\ & \ddots & \\ & & \lambda_{d-1}^{1/2} \end{pmatrix} U^*.$$

Then immediately B is a positive-semidefinite matrix with $B^2 = A$. It is also possible to prove that B is the only positive semidefinite square root.

Absolute value of a matrix

Now for an arbitrary matrix A, define:

$$|A| = \sqrt{A^* A}.$$

The matrix $|A|$ is the quantum analogue of the absolute value of a complex number (as the notation suggests). The absolute value can also be counterintuitive:

Exercise 1.5

All of the following assertions are false in general:

(a) $|A + B| \leq |A| + |B|$,

(b) $|AB| \leq |A||B|$,

(c) $|A| = |A^*|$.

Find counterexamples for each of them.

Inadequacies of the absolute value

The absolute value $|A|$ of a matrix A is the correct generalisation of the numerical absolute value in the quantised case, but the above counterintuitive properties make is unsuitable for studying problems in analysis such as convergence of infinite series.
This is one motivation to introduce the notion of a matrix norm.

Definition 1.3

A matrix norm in dimension $d \geq 1$ is a mapping $\| \cdot \| : M_d \to [0, \infty)$ which satisfies the following conditions:

(a) $\|A\| = 0$ if and only if $A = 0$;

(b) $\|\lambda A\| = |\lambda| \, \|A\|$ for all $A \in M_d$ and $\lambda \in \mathbb{C}$ (Positive homogeneity)

(c) $\|A + B\| \leq \|A\| + \|B\|$ for all $A, B \in M_d$ (Triangle inequality).

The operator norm

The most important matrix norm is the operator norm:

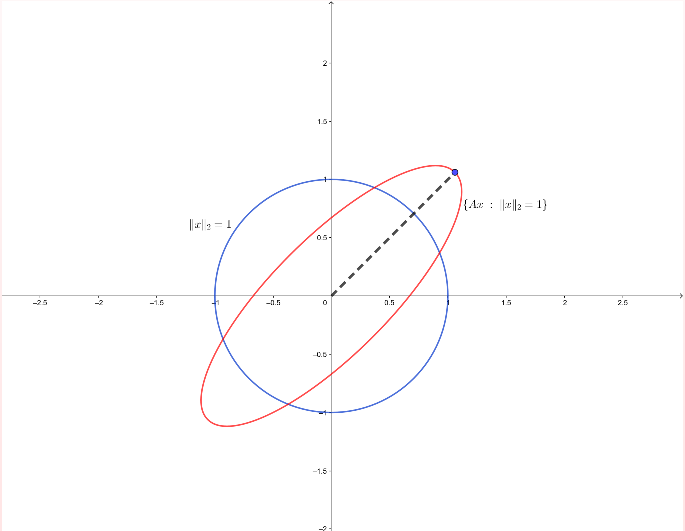
$$\|A\|_{\mathrm{op}} = \sup_{x \in \mathbb{C}^d \setminus \{0\}} \frac{\|Ax\|_2}{\|x\|_2}$$

where $\|\cdot\|_2$ is the Euclidean norm, $\|x\|_2 = \left(\sum_{j=0}^{d-1} |x_j|^2\right)^{1/2}$.
Equivalently (by positive homogeneity),

$$\|A\|_{\mathrm{op}} = \sup_{\|x\|_2 = 1} \|Ax\|_2.$$

Visualising $\|A\|_{\mathrm{op}}$ in $\mathbb{R}^2$

Here, the blue circle is the set $\{x \in \mathbb{R}^2 \ : \ \|x\|_2 = 1\}$, and A is a matrix that moves the blue circle to the red ellipse. The maximum norm on the ellipse is the operator norm $\|A\|_{\mathrm{op}}$.



Exercise 1.6

Verify that $\|\cdot\|_{\mathrm{op}}$ is a matrix norm.

Lemma 1.4

Let $A \in M_d$ and $x \in \mathbb{C}^d$. Then

$$\|Ax\|_2 \leq \|A\|_{\mathrm{op}}\|x\|_2$$

Proof.

By definition, for any $x \in \mathbb{C}^d \neq \{0\}$:

$$\|A\|_{\mathrm{op}} \geq \frac{\|Ax\|_2}{\|x\|_2}.$$

$\square$

The following result is called the submultiplicativity of the operator norm:

Lemma 1.5

Let $A, B \in M_d$. Then:

$$\|AB\|_{\mathrm{op}} \leq \|A\|_{\mathrm{op}} \|B\|_{\mathrm{op}}.$$

Proof.

Let $x \in \mathbb{C}^d$. Then,

$$\|ABx\|_2 \leq \|A\|_{\mathrm{op}} \|Bx\|_2 \leq \|A\|_{\mathrm{op}} \|B\|_{\mathrm{op}} \|x\|_2.$$

So,

$$\sup_{\|x\|_2 = 1} \|ABx\|_2 \leq \|A\|_{\mathrm{op}} \|B\|_{\mathrm{op}}.$$

$\square$

By induction, for all $n \geq 1$ we have:

$$\|A^n\|_{\mathrm{op}} \leq \|A\|_{\mathrm{op}}^n.$$

It is this inequality which allows us to answer the questions at the beginning of the lecture.

Other matrix norms

The operator norm is not the only useful norm on M_d! Another very important norm is the Hilbert-Schmidt norm:

$$\|A\|_{\mathcal{L}_2} = \left(\sum_{0 \le j,k \le d-1} |\langle e_j, Ae_k \rangle|^2 \right)^{1/2} .$$

This norm has various names: it is also called the Frobenius norm or the Schatten-von Neumann $\mathcal{L}_2$ norm, or just $\mathcal{L}_2$).
Viewed another way, the $\mathcal{L}_2$-norm is the Euclidean norm of A considered as a point of $\mathbb{C}^{d^2}$.

Exercise 1.7
Verify that $\|\cdot\|_{\mathcal{L}_2}$ is a matrix norm.

Convergence of matrices

Let $\{A_k\}_{k=0}^{\infty}$ be a sequence in $M_d(\mathbb{C})$. There are many senses in which we can say that $\lim_{k\to\infty} A_k = A$, for some $A \in M_d$.

Definition 1.6

The sequence $\{A_k\}_{k=0}^{\infty}$ is said to converge in the sense of matrix elements to $A \in M_d$ if for all $0 \le j, l \le d-1$ we have:

$$\lim_{k\to\infty} \langle e_j, A_k e_l \rangle = \langle e_j, A e_l \rangle.$$

(Recall, e_j denotes the jth standard basis vector).

Definition 1.7

Let $\|\cdot\|$ be any matrix norm. The sequence $\{A_k\}_{k=0}^{\infty}$ is said to converge in the $\|\cdot\|$ sense if we have:

$$\lim_{k\to\infty} \|A_k - A\| = 0.$$

Convergence of matrices

Despite looking apparently different, convergence in norm and convergence in entries are equivalent.

Lemma 1.8

We have that $A_k \to A$ in the sense of matrix elements $\iff$ $A_k \to A$ in the sense of a matrix norm.

Proof.

Exercise.
(Hint: for the $\Rightarrow$ direction use the triangle inequality. For the $\Leftarrow$ direction use the fact that closed and bounded subsets of $\mathbb{C}^{d^2}$ are compact, and continuous functions on compact sets are bounded.) $\square$

Warning: These notions of convergence are equivalent only in finite dimensions. Later in the course we will study spaces of infinite matrices which will have multiple inequivalent notions of convergence.

Convergence of Cauchy sequences

A sequence $\{A_k\}_{k\to\infty}$ is called Cauchy in the norm $\|\cdot\|$ if for all $\varepsilon > 0$, there exists $K > 0$ such that for all $k, n > K$ we have:

$$\|A_k - A_n\| < \varepsilon.$$

Theorem 1.9

If $\{A_k\}_{k=0}^{\infty}$ is a Cauchy sequence in any matrix norm, then A_k has a limit.

Proof.

Exercise.

(Hint: Using the equivalence of matrix norm convergence and matrix element convergence, show that the sequences of entries of $\{A_k\}_{k=0}^{\infty}$ are Cauchy in $\mathbb{C}$. Then use the fact that every Cauchy sequence in $\mathbb{C}$ has a limit to show that a limit $A \in M_d$ exists). $\square$

Convergence of infinite series

Proposition 1.10

Let $A \in M_d$. There exists a unique matrix $\exp(A)$ *such that:*

$$\exp(A) = \lim_{n\to\infty} \sum_{k=0}^{n} \frac{A^k}{k!}$$

(where $A^0 = 1_d$, including $0^0 = 1_d$). The convergence here can be taken in the sense of matrix elements.

Convergence of infinite series

Let

$$s_n = \sum_{k=0}^{n} \frac{A^k}{k!}, \quad n \geq 0.$$

For any $n > m \geq 0$, we have:

$$\|s_n - s_m\|_{\mathrm{op}} \leq \sum_{k=m+1}^{n} \frac{\|A\|_{\mathrm{op}}^k}{k!} \leq \exp(\|A\|_{\mathrm{op}}) - \sum_{k=0}^{m} \frac{\|A\|_{\mathrm{op}}^k}{k!}.$$

Since $\lim_{m\to\infty} \sum_{k=0}^{m} \frac{\|A\|_{\mathrm{op}}^k}{k!} = \exp(\|A\|_{\mathrm{op}})$, it follows that the sequence $\{s_n\}_{n\geq 0}$ is Cauchy.$\square$

Convergence of infinite series

Let $\| \cdot \|$ be any matrix norm on M_d.

Exercise 1.8

Show that if $\|A\| < 1$, then:

$$(1_d - A)^{-1} = \sum_{j=0}^{\infty} A^j$$

where the series converges in the sense of matrix elements, or equivalently in the norm sense.

Exercise 1.9

Let $f(z) = \sum_{j=0}^{\infty} c_j z^j$ be any function holomorphic in a neighbourhood of $z = 0$. Show that if $\|A\|_{\mathrm{op}}$ is less than the radius of convergence of f at 0 then the infinite series $\sum_{j=0}^{\infty} c_j A^j$ converges.

As before, we have $A^0 = 1_d$ for every matrix A including $A = 0$.

Matrices as quantised numbers

From the "quantised numbers" perspective, the matrix exponential $A \mapsto \exp(A)$ is the quantum analogue of the classical exponential $t \mapsto \exp(t)$.

The trace

Let $A \in M_d$. The trace of A is the sum of its diagonal entries:

$$\mathrm{Tr}(A) = \sum_{j=0}^{d-1} A_{j,j}.$$

Exercise 1.10

Check that for all $A, B \in M_d$ we have:

$$\mathrm{Tr}(AB) = \mathrm{Tr}(BA).$$

While you're at it: check that the same identity holds with non-square matrices, $A \in M_{n,m}(\mathbb{C})$ and $B \in M_{m,n}(\mathbb{C})$.

We can further say that:

$$\mathrm{Tr}(A_1 A_2 \cdots A_k) = \mathrm{Tr}(A_2 \cdots A_k A_1).$$

This property is called the cyclicity of the trace.

Exercise 1.11

Show that if $A \geq 0$ then $\mathrm{Tr}(A) \geq 0$.

The importance of the trace lies in the fact that $\mathrm{Tr}(A)$ is independent of the choice of coordinate system.
Indeed, from cyclicity, if B is an invertible matrix then:

$$\mathrm{Tr}(A) = \mathrm{Tr}(B^{-1}AB), \quad A \in M_d.$$

Corollary: if $\{e_0, \ldots, e_{d-1}\}$ and $\{f_0, \ldots, f_{d-1}\}$ are any two bases for $\mathbb{C}^d$, then:

$$\sum_{j=0}^{d-1} \langle e_j, Ae_j \rangle = \sum_{j=0}^{d-1} \langle f_j, Af_j \rangle = \mathrm{Tr}(A).$$

In fact, Tr is essentially the only coordinate independent linear functional on M_d:

Exercise 1.12

Show that if $\theta : M_d(\mathbb{C}) \to \mathbb{C}$ is linear, and has the property that $\theta(A) = \theta(U^* AU)$ for all unitary matrices U and $A \in M_d$, then θ is a scalar multiple of the trace: $\theta = \lambda \mathrm{Tr}$,

Trace as expectation value

If we have a matrix $A \in M_d(\mathbb{C})$, then the normalised trace $\frac{1}{d}\mathrm{Tr}(A)$ is the average possible value:

$$\frac{1}{d}\mathrm{Tr}(A) = \frac{1}{d}\sum_{j=0}^{d-1}\lambda_j(A).$$

Remark 1

In quantum mechanics, self-adjoint operators (and in particular matrices) are called observables. The spectrum of an observable (i.e., a matrix A) is $\{\lambda_0(A), \lambda_1(A), \ldots, \lambda_{d-1}(A)\}$, *and is the range of possible observed values of A.*

Recall that the adjoint A^* of a matrix is its conjugate transpose: $(A^*)_{j,k} = \overline{A_{k,j}}$.

Exercise 1.13

Show that for all $A, B \in M_d$:

$$\mathrm{Tr}(B^* A) = \sum_{0 \le j,k \le d-1} \overline{B}_{j,k} A_{j,k}.$$

So in fact:

$$\|A\|_{\mathcal{L}_2} = \mathrm{Tr}(A^* A)^{1/2}.$$

Trace and eigenvalues

Another perspective relates the trace and the determinant:

$$\det(1_d + \lambda A) = 1 + \mathrm{Tr}(A)\lambda + O(\lambda^2), \quad \lambda \to 0.$$

meaning that the coefficient of λ in the polynomial $\det(1_d + \lambda A)$ is $\mathrm{Tr}(A)$. (Exercise: prove this.)

It then follows from the algebraic properties of the determinant that:

$$\det(A - \lambda 1_d) = (-\lambda)^d + \mathrm{Tr}(A)(-\lambda)^{d-1} + O(\lambda^{d-2}).$$

Recall that the eigenvalues $\{\lambda_0(A), \ldots, \lambda_{d-1}(A)\}$ are the roots of the polynomial $\det(A - \lambda 1_d)$.
So by the Newton identities:

$$\sum_{j=0}^{d-1} \lambda_j(A) = \mathrm{Tr}(A).$$

The "magic" of the above formula is that computing each eigenvalue $\lambda_j(A)$ individually is hard, but computing their sum is trivial (just sum up the diagonal entries of A).