
Matrix Analysis for Quantum Information Theory



Institute for
Mathematical Sciences

Koenraad M.R. Audenaert



Quantum Information at
Imperial College
London



Introduction

- Books on Matrix Analysis:
 - R. Bhatia, Matrix Analysis, Springer, 1997.
 - R.A. Horn and C.R. Johnson, Matrix Analysis, Cambridge, 1985.
 - R.A. Horn and C.R. Johnson, Topics in Matrix Analysis, Cambridge, 1991.
 - X. Zhan, Matrix Inequalities, Springer, 1999.
- Journals: Linear Algebra and its Applications, and others
- Purpose of Lectures: introduction to those aspects of matrix analysis that are/may be/have been useful in QIT
- Connections with QIT
- No Proofs
- Background: maths used in QM



Contents (roughly)

- Classes of matrices
- Operations and functions on matrices
- Decompositions
- Orderings
- Norms
- Inequalities



Matrices

- A matrix represents a linear transformation on a vector space: $f(x)$ represented by $A(f).x$
- Matrix = square or rectangular array of... real or complex numbers, “entries”;
 $A = (A_{i,j})_{i=1..n,j=1..m}$
- ... or of matrices: block matrices
- Concatenation of transformations: matrix product; $(g \circ f)(x) = A(g).A(f).x$
- Matrix product is non-commutative: $AB \neq BA$
- Examples in QIT: occur whenever systems are finite-dimensional (in one way or another); density matrix of a state, observables, Hamiltonians



Matrix Operations 1

- Inverse: A^{-1} , satisfies $AA^{-1} = \mathbf{I}$; need not always exist
- Transpose: A^T , $(A^T)_{i,j} = A_{j,i}$
- Complex Conjugate: \bar{A} , $\bar{A}_{i,j} = \overline{A_{i,j}}$
- Hermitian Conjugate: $A^* = \overline{A^T}$; in physics: A^\dagger
- Example: if $A = |\psi\rangle$, then $A^* = \langle\psi|$
- $(AB)^T = B^T A^T$
- $\overline{AB} = \bar{A} \bar{B}$
- $(AB)^* = B^* A^*$



Matrix Classes

- Diagonal matrix: square matrix with non-zero elements on diagonal only:
 $A_{i,j} = a_i \delta_{i,j}$ or $A = \text{Diag}(a_1, a_2, \dots)$
- Identity matrix \mathbf{I} : diagonal matrix with all 1's on the diagonal: $\mathbf{I}_{i,j} = \delta_{i,j}$
- Scalar matrix: $A = a\mathbf{I}$
- Normal matrix: A commutes with A^* , i.e. $AA^* = A^*A$
- Hermitian matrix: $A = A^*$
- Positive semi-definite (PSD) matrix: $\exists B : A = B^*B$; denoted $A \geq 0$
- Unitary matrix: square matrix U with $U^*U = \mathbf{I}$



Matrix Classes

- Symmetric matrix: $A = A^T$; = Hermitian when real
- Orthogonal matrix: square matrix O with $O^T O = \mathbf{I}$; = unitary when real
- Skew-Hermitian: $A^* = -A$; if B is Hermitian, $A = iB$ is skew-Hermitian
- Skew-symmetric: $A^T = -A$



Characterisations 1

- Examples of Hermitian matrices: observables, Hamiltonians
- Example of PSD matrices: density matrices; e.g. $A = |\psi\rangle\langle\psi|$; $B^* = |\psi\rangle$
- Examples of unitary matrices: any evolution operator, Pauli matrices, CNOT
- A matrix A is Hermitian iff all its expectation values are real:
 $\langle\psi|A|\psi\rangle \in \mathbb{R}$
- A matrix A is PSD iff all its expectation values are real and non-negative:
 $\langle\psi|A|\psi\rangle \geq 0$, whence the name
- Exercise: prove this.
- A matrix is unitary iff its column vectors form an orthonormal basis
- For square U , $U^*U = \mathbf{I}$ implies $UU^* = \mathbf{I}$



Matrix Operations 2

- A new matrix product: the tensor, or Kronecker product $A_1 \otimes A_2$
- Example: density matrix of “independent” particles: product state
- Can be thought of as an ordered list (A_1, A_2) with the following rules:
 1. Product: $(A_1, A_2) \cdot (B_1, B_2) = (A_1 B_1, A_2 B_2)$
 2. Reduces to ordinary product when all A_i are scalars a_i
- It can be represented by a block matrix:

$$(A \otimes B)_{(i,j),(k,l)} = A_{ik} B_{jl},$$

where (i, j) is a *composite* (row) index: i, k index the blocks, and j, l index within the blocks.

- E.g. when A is 2×2

$$A \otimes B = \begin{pmatrix} A_{11}B & A_{12}B \\ A_{21}B & A_{22}B \end{pmatrix}$$



Matrix Operations 3

- For square matrices: trace, determinant
- Trace: $\text{Tr}(A) = \sum_i A_{i,i}$
 - Trace is a linear operation
 - Cyclicity property: $\text{Tr}(AB) = \text{Tr}(BA)$, $\text{Tr}(ABC) = \text{Tr}(BCA)$,...
 - $\text{Tr}(A \otimes B) = \text{Tr}(A) \text{Tr}(B)$
- Determinant: $\det(A)$ (I hope you know the definition)
 - Det of an $n \times n$ matrix is homogeneous of order n : $\det(aA) = a^n \det(A)$
 - For square matrices A, B, C, \dots : $\det(ABC\dots) = \det(A) \det(B) \det(C)\dots$



Matrix Functions

- Analytic functions can be represented by (formal) power series $f(z) = \sum_{k=0} a_k z^k$
- Since we know how to multiply matrices we can calculate $\sum_{k=0} a_k A^k$
- This (formally) defines a matrix function $f(A)$
- Example: $\exp(A) = \sum_{k=0} A^k / k!$



Matrix Functions

- Other functions may be defined as an integral; this also carries over to matrices
- Example: x^p for $0 < p \leq 1$ and $x > 0$:

$$x^p = \frac{\sin(p\pi)}{p} \int_0^\infty \frac{\lambda x}{\lambda + x} \lambda^{p-2} d\lambda$$

- Then, for a PSD matrix A :

$$A^p = \frac{\sin(p\pi)}{p} \int_0^\infty \lambda A(\lambda \mathbf{I} + A)^{-1} \lambda^{p-2} d\lambda$$

- Also, for a PSD matrix A :

$$\log(\mathbf{I} + A) = \int_1^\infty \lambda A(\lambda \mathbf{I} + A)^{-1} \lambda^{-2} d\lambda.$$

- Fortunately, there are easier ways, which we will see below!



Matrix Operations 4

- Cartesian decomposition:

any square matrix T can be written as $T = A + iB$, with A and B Hermitian:

$$A = (T + T^*)/2, \quad B = (T - T^*)/(2i)$$

A = Hermitian part, B = skew-Hermitian part

- Matrix absolute value (or modulus): $|A| = (A^*A)^{1/2}$
- Jordan decomposition: any Hermitian matrix H can be written as $H = H^+ - H^-$, with H^+ and H^- PSD, and $H^+H^- = 0$:

$$H^+ = (|H| + H)/2, \quad H^- = (|H| - H)/2, \quad |H| = H^+ + H^-$$

H^+ = positive part, H^- = negative part



Eigenvalues

- Many of the presented concepts will get “easier” descriptions when the matrix has an eigenvalue decomposition.
- Eigenvalue/eigenvector: $Ax = \lambda x$
- Stack eigenvectors $x^{(i)}$ columnwise in matrix S , eigenvalues λ_i in diagonal matrix Λ :
 $AS = S\Lambda$
- If S is invertible, we get $A = S\Lambda S^{-1}$
- A matrix is *diagonalisable* if there exists an invertible S such that $S^{-1}AS$ is diagonal.
- A matrix is *unitarily diagonalisable* if there exists a unitary U such that $U^{-1}AU = U^*AU$ is diagonal; then $A = U\Lambda U^*$.



Eigenvalues

- Theorem: A matrix is unitarily diagonalisable (UD) iff the matrix is normal
- Exercise: prove \Rightarrow part
- The eigenvalue decomposition (EVD) of a normal matrix A is $A = U\Lambda U^*$
- A Hermitian matrix is UD, with real eigenvalues
- A PSD matrix is UD, with non-negative eigenvalues
- EVD of a real symmetric matrix: $A = O\Lambda O^T$, with O real orthogonal
- Exercise: For Hermitian H , express $\text{Tr}(H)$, $|H|$, H^+ , H^- in terms of its EVD
- Matrix functions of Hermitian (or PSD) matrices: $f(A) = Uf(\Lambda)U^*$, where f operates entrywise on the diagonal elements (eigenvalues)
- Example: for PSD A , with $A = U\Lambda U^*$

$$A^{1/2} = U \text{Diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \dots)U^*$$



Eigenvalues

- If A and B are square, AB and BA have the same set of eigenvalues
- If A (or B) is invertible, there is a simple proof
- If A and B are non-square, AB and BA have the same set of non-zero eigenvalues
- If A and B are PSD, AB has non-negative real eigenvalues (see below)
- If eigenvalues of A and B are λ_i and μ_j , eigenvalues of $A \otimes B$ are $\lambda_i \mu_j$



Exercise

- Consider a finite-dimensional quantum system in a pure state $\psi(t)$, evolving under a time-independent Hamiltonian H
- Calculate the evolution operator $U(t)$: $\psi(0) \mapsto \psi(t) = U(t)\psi(0)$
- Show that the evolution operator is unitary. What is the physical necessity for $U(t)$ to be unitary (hint: what quantity has to be preserved)?
- Express $U(t)$ in terms of the EVD of H , given by $H = V\Lambda V^*$, with V unitary.



Singular values

- Not all matrices are diagonalisable; example: Jordan block; e.g.

$$\begin{pmatrix} x & 1 \\ 0 & x \end{pmatrix}$$

- However, all matrices, even the non-square ones, have a singular value decomposition (SVD): $A = U\Sigma V^*$, where U and V are unitary and Σ is “diagonal”.
- One can find U and V s.t. the diagonal elements of Σ are non-negative reals and sorted in non-ascending fashion; then the diagonal elements of Σ , $\sigma_i(A)$, are the singular values of A .
- Exercise: $\sigma_i(A) = \lambda_i(|A|)$
- For real A , U and V can be taken to be real orthogonal



Singular values and Invertibility

- One of the ways to check invertibility of a square matrix is to inspect its singular values: A is invertible iff all $\sigma_i(A) > 0$, strictly.
- An easier way is just to calculate $\det(A)$: A is invertible iff $\det(A) \neq 0$.
- Exercise: Prove this using the SVD of A . What can one say about $\det(U)$ for a unitary matrix U ?
- The SVD reveals more than just invertibility: the number of non-zero singular values equals the *rank* of A = the number of independent column (or row) vectors of A .



Schmidt decomposition

- Consider a pure bipartite state $\psi = \sum_{i=1..d_A, j=1..d_B} x_{i,j} |i\rangle_A |j\rangle_B$
- Schmidt decomposition: one can find orthonormal bases u_i and v_j of the A and B systems, respectively, so that $\psi = \sum_i c_i |u_i\rangle_A |v_i\rangle_B$, where the Schmidt coefficients c_i are non-negative.
- Tilde notation: denote the matrix of coefficients $x_{i,j}$ by $\tilde{\psi}$
- The Schmidt decomposition of ψ is the SVD of the matrix $\tilde{\psi}$.
Nothing new under the Sun!
- Normalisation: $1 = \langle \psi | \psi \rangle = \text{Tr}(\tilde{\psi} \tilde{\psi}^*) = \sum_i c_i^2$.



Orderings, and their preservers

- Infinitely many ways of imposing a (partial) order on a class of matrices
- Of particular importance are: the majorisation order, the PSD order, and norm orderings
- We will discuss these orderings, and also the matrix operations that preserve them



Majorisation

- The majorisation order is basically defined for real vectors (Bhatia, Chapter II)
- Notation (Bhatia): let a, b be n -dimensional real vectors; b majorises $a = a \prec b$ iff

$$\sum_{i=1}^k a_i^\downarrow \leq \sum_{i=1}^k b_i^\downarrow, \quad 1 \leq k \leq n$$

with equality for $k = n$; else it is weak majorisation, $a \prec_w b$.

- Its definition can be extended to matrices via the singular values: $A \prec B$ iff

$$\sum_{i=1}^k \sigma_i(A) \leq \sum_{i=1}^k \sigma_i(B), \quad 1 \leq k \leq n$$

with equality for $k = n$.



Majorisation

- Example: for probability vectors x (non-negative real vectors with $\sum_i x_i = 1$):

$$\left(\frac{1}{n}, \dots, \frac{1}{n}\right) \prec x \prec (1, 0, \dots, 0)$$

- Many interesting relations can be expressed in terms of majorisation
- Schur's Theorem: for any Hermitian matrix A , $\text{Diag}(A) \prec \lambda(A)$
- For Hermitian A and B , $\lambda^\downarrow(A+B) \prec \lambda^\downarrow(A) + \lambda^\downarrow(B)$
- For any A and B , $\sigma(A+B) \prec_w \sigma(A) + \sigma(B)$
- For any A and B , and $r > 0$ $\sigma^r(AB) \prec_w \sigma^r(A)\sigma^r(B)$



Majorisation in QIT

- Nielsen's Theorem: a pure bipartite state ψ with Schmidt coefficients a can be LOCC-converted to a pure state ϕ with Schmidt coefficients b iff $a \succ b$; that is, iff $\tilde{\psi} \succ \tilde{\phi}$.



Majorisation preservers

- $f : \mathbb{R}^n \mapsto \mathbb{R}$ is Schur-convex iff $x \prec y \Rightarrow f(x) \leq f(y)$
- $\Phi : \mathbb{R}^n \mapsto \mathbb{R}^m$ is isotone iff $x \prec y \Rightarrow \Phi(x) \prec_w \Phi(y)$
- Φ is strongly isotone iff $x \prec_w y \Rightarrow \Phi(x) \prec_w \Phi(y)$
- Φ is strictly isotone iff $x \prec y \Rightarrow \Phi(x) \prec \Phi(y)$
- Example of Schur-convex function: $f(x) = -\prod_j x_j$ on \mathbb{R}_+^n
- Examples of isotone maps: absolute value, square
- Examples of strongly isotone maps: powers $p > 1$, on \mathbb{R}_+^n
- If $t \mapsto f(e^t)$ is convex and monotonely increasing then
 $x \prec_{\log} y \equiv \log x \prec_w \log y \Rightarrow f(x) \prec_w f(y)$



The PSD ordering

- PSD matrix: $A \geq 0$.
- For Hermitian A, B , $A \geq B$ iff $A - B \geq 0$: PSD ordering
- Linear operations that preserve PSD ordering: positive maps, including the completely positive maps; example: $A \mapsto XAX^*$
- Appl.: for $A, B \geq 0$, $\lambda(AB) = \lambda(AC^*C) = \lambda(CAC^*) \geq 0$
- Matrix functions that preserve PSD ordering: matrix monotones
- Examples: powers p with $0 < p \leq 1$, \log
- Powers with $p > 1$ are NOT matrix monotone:

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, B = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}, A \leq B, A^2 \not\leq B^2$$



Matrix Norms

- A matrix norm $\|A\|$ is a mapping from the space of matrices to \mathbb{R}_+ obeying:
 - $\|A\| = 0$ iff $A = 0$
 - Homogeneous: $\|zA\| = |z| \|A\|$
 - Triangle inequality: $\|A + B\| \leq \|A\| + \|B\|$
 - Submultiplicative: $\|AB\| \leq \|A\| \|B\|$
- Exercise: prove that they are convex functions
- Of particular interest are the unitarily invariant (UI) matrix norms: $\|UAV\| = \|A\|$, i.e. they depend only on $\sigma(A)$



UI Matrix Norms

- Operator norm: $\|A\| = \sigma_1(A)$, largest singular value
- Trace norm: $\|A\|_{\text{Tr}} = \|A\|_1 = \sum_{i=1}^n \sigma_i(A) = \text{Tr} |A|$
- Frobenius or Hilbert-Schmidt norm:
$$\|A\|_2 = \left(\sum_{i=1}^n \sigma_i^2(A) \right)^{1/2} = \left(\text{Tr} |A|^2 \right)^{1/2} = \left(\sum_{i,j} |A_{i,j}|^2 \right)^{1/2}$$
- Schatten q -norms: $\|A\|_q = \left(\sum_{i=1}^n \sigma_i^q(A) \right)^{1/q} = \left(\text{Tr} |A|^q \right)^{1/q}$
- Ky Fan norms: $\|A\|_{(k)} = \sum_{i=1}^k \sigma_i(A)$
- Ky Fan Dominance Theorem:
 $\|A\| \leq \|B\|$ for all UI norms, iff this holds for all Ky Fan norms
- Exercise: find, in these notes, the statement that the Ky Fan norms satisfy the triangle inequality and submultiplicativity



Matrix norms in QIT

- Matrix norms are important in QIT for many reasons
- To express that two sequences of states ρ_n, σ_n are asymptotically indistinguishable: $\lim_{n \rightarrow \infty} \|\rho_n - \sigma_n\| = 0$
- Any UI norm of $\rho - \sigma$ can be used as a distance measure between states
- Schatten q -norm of a state is a measure of its purity
- The von Neumann entropy $S(\rho) = -\text{Tr}(\rho \log \rho)$ has two friends:
the Tsallis entropies $S_\alpha(\rho) = \text{Tr} \rho^\alpha = \|\rho\|_\alpha^\alpha$,
and the Renyi entropies $\log(\text{Tr} \rho^\alpha)/(1 - \alpha)$



Norm inequalities

- Cauchy-Schwarz inequality: $|||A^*XB|||^2 \leq |||AA^*X||| \ |||XBB^*|||$
- Arithmetic-Geometric inequality: $|||A^*XB||| \leq |||(AA^*X + XBB^*)/2|||$
- Hölder's inequality: for $p, q \geq 1$ s.t. $1/p + 1/q = 1/r$,
 $||| |AB|^r |||^{1/r} \leq ||| |A|^p |||^{1/p} \ ||| |B|^q |||^{1/q}$
- Araki-Lieb-Thirring: for $A, B \geq 0$, and $0 \leq t \leq 1$, $|||B^t A^t B^t||| \leq |||(BAB)^t|||$,
while the inequality is reversed for $t \geq 1$



The Chernoff Bound

From “Distinguishability and Accessible Information in Quantum Theory,”
PhD thesis of Christopher A. Fuchs (then at UNM):

Chernoff distance between two distributions p and q :

$$Q(p, q) = \min_{0 \leq s \leq 1} \sum_i p_i^s q_i^{1-s}$$

Error exponent in symmetric hypothesis test for discriminating between p and q
given by $-\log Q(p, q)$.



The Chernoff Bound

Consider coloured balls in an urn

- Hypothesis H_0 : colours distributed according to p
- Hypothesis H_1 : colours distributed according to q

How many times (n) do we have to draw before we can tell which hypothesis is true?

Total error probability = sum of type I and type II error probabilities

Goes to 0 exponentially with n , at a rate $-\log Q(p, q)$ [Chernoff '52].



The Quantum Smirnov Bound

Question: What should the quantum version be of this Chernoff Bound? What is the error exponent in the symmetric hypothesis test for discriminating between two *quantum states* ρ and σ ?

- Hypothesis H_0 : n draws yield state $\rho^{\otimes n}$
- Hypothesis H_1 : n draws yield state $\sigma^{\otimes n}$

Quantum measurement theory by Helström and Holevo from 70's:
the total error probability of the optimal measurement scheme is

$$P_e = (1 - \|\rho^{\otimes n} - \sigma^{\otimes n}\|_1/2)/2.$$



The Quantum Chernoff Bound

The P_e goes down exponentially with n at the rate

$$\lim_{n \rightarrow \infty} -\frac{1}{n} \log(1 - \|\rho^{\otimes n} - \sigma^{\otimes n}\|_1/2)$$

Finding a closed-form expression for the error rate is an open problem in QIT!

Partial answers:

- Ogawa and Hayashi (2004): three candidate expressions
candidate #2:

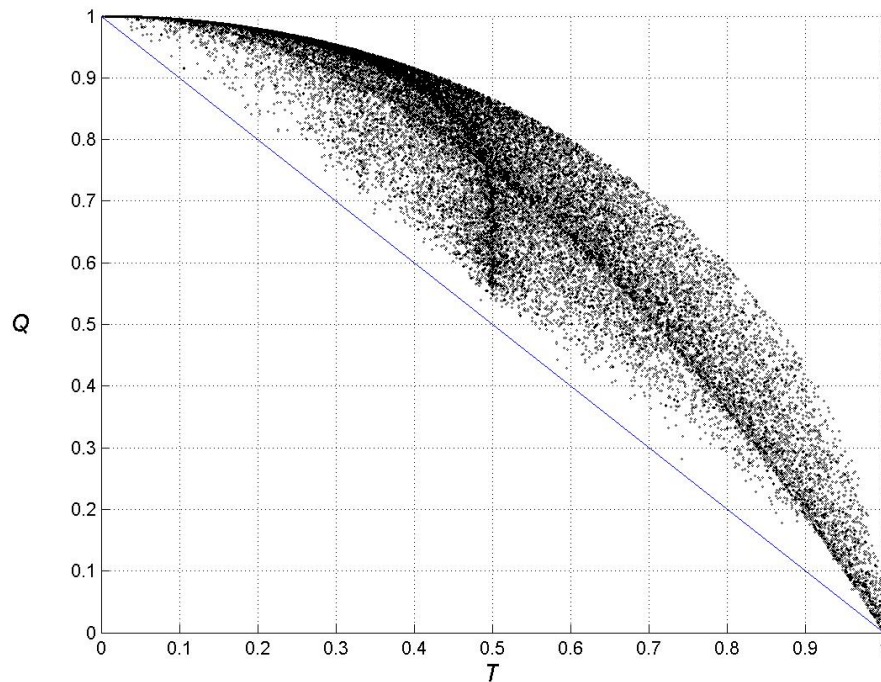
$$-\log \left(\min_{0 \leq s \leq 1} \text{Tr}[\rho^s \sigma^{1-s}] \right)$$

- Nussbaum and Szkola (July 2006): candidate #2 is an upper bound



Matlab calculations

Scatter plot of $Q(\rho, \sigma) := \min_{0 \leq s \leq 1} \text{Tr}[\rho^s \sigma^{1-s}]$ versus trace norm distance $T(\rho, \sigma) := \|\rho - \sigma\|_1/2$.





Matlab calculations

From the scatter plot we get the impression that for any pair of states one has

$$Q + T \geq 1.$$

That would in fact be enough to prove that Hayashi's candidate #2 is not only an upper bound but also a lower bound on the error exponent, hence equal to it.

Basic reason for that is the multiplicativity of Q :

$$\frac{1}{n} \log Q(\rho^{\otimes n}, \sigma^{\otimes n}) = \log Q(\rho, \sigma)$$

This would solve a long-standing open problem in QIT!



Matrix Analysis yields the answer

More generally, we have

Theorem 1 *Let A and B be two finite-dimensional, positive semi-definite matrices. Then, for all $s \in [0, 1]$ one has*

$$P_e(A, B) := \operatorname{Tr} |A - B| + 2 \operatorname{Tr} A^s B^{1-s} - \operatorname{Tr} A - \operatorname{Tr} B \geq 0.$$

This follows from (Audenaert et al, quant-ph/0610027):

Lemma 1 *Let $A, B \geq 0$. Let $0 \leq t \leq 1$, and let P be the projector on the range of the positive part of $A - B$. Then*

$$\operatorname{Tr}[PB(A^t - B^t)] \geq 0.$$