

Contents

A	Linear algebra basics	1
A.1	Basic operations	1
A.2	Fundamental theorem of linear algebra	4
A.2.1	Part one	4
A.2.2	Part two	5
A.3	Nature of the solution	7
A.3.1	Exactly-identified	7
A.3.2	Under-identified	8
A.3.3	Over-identified	10
A.4	Matrix decomposition and inverse operations	12
A.4.1	LU factorization	12
A.4.2	Cholesky decomposition	16
A.4.3	Eigenvalues and eigenvectors	17
A.4.4	Singular value decomposition	23
A.4.5	Spectral decomposition	28
A.4.6	quadratic forms, eigenvalues, and positive definiteness	29
A.4.7	similar matrices, Jordan form, and generalized eigen- vectors	29
A.5	Gram-Schmidt construction of an orthogonal matrix	32
A.5.1	QR decomposition	34
A.5.2	Gram-Schmidt QR algorithm	34
A.5.3	Accounting example	35
A.5.4	The Householder QR algorithm	36
A.5.5	Accounting example	36

A.6	Computing eigenvalues	39
A.6.1	Schur's lemma	39
A.6.2	Power algorithm	40
A.6.3	QR algorithm	41
A.6.4	Schur decomposition	43
A.7	Some determinant identities	46
A.7.1	Determinant of a square matrix	46
A.7.2	Identities	47
A.8	Matrix exponentials and logarithms	49
B	Iterated expectations	51
B.1	Decomposition of variance	53
B.2	Jensen's inequality	54
C	Multivariate normal theory	55
C.1	Conditional distribution	57
C.2	Special case of precision	61
C.3	Truncated normal distribution	63
D	Projections and conditional expectations	71
D.1	Gauss-Markov theorem	71
D.2	Generalized least squares (GLS)	74
D.3	Recursive least squares	76
E	Two stage least squares IV (2SLS-IV)	79
E.1	General case	79
E.2	Special case	81
F	Seemingly unrelated regression (SUR)	83
F.1	Classical	84
F.2	Bayesian	84
F.3	Bayesian treatment effect application	85
G	Maximum likelihood estimation of discrete choice models	87
H	Optimization	89
H.1	Linear programming	89
H.1.1	basic solutions or extreme points	90
H.1.2	fundamental theorem of linear programming	91
H.1.3	duality theorems	91
H.1.4	example	92
H.1.5	complementary slackness	92
H.2	Nonlinear programming	93
H.2.1	unconstrained	93
H.2.2	convexity and global minima	93
H.2.3	example	94

H.2.4	constrained — the Lagrangian	94
H.2.5	Karash-Kuhn-Tucker conditions	95
H.2.6	example	96
H.3	Theorem of the separating hyperplane	97
I	Quantum information	99
I.1	Quantum information axioms	99
I.1.1	The superposition axiom	99
I.1.2	The transformation axiom	100
I.1.3	The measurement axiom	100
I.1.4	The combination axiom	101
I.2	Summary of quantum "rules"	103
I.3	Observables and expected payoffs	104
I.4	Density operators and quantum entropy	105
I.4.1	Quantum entropy	108
I.5	Some trigonometric identities	108
J	Common distributions	111

Appendix I

Quantum information

Quantum information follows from physical theory and experiments. Unlike classical information which is framed in the language and algebra of set theory, quantum information is framed in the language and algebra of vector spaces. To begin to appreciate the difference, consider a set versus a tuple (or vector). For example, the tuple $[1\ 1]$ is different than the tuple $[1\ 1\ 1]$ but the sets, $\{1, 1\}$ and $\{1, 1, 1\}$ are the same as set $\{1\}$.

Quantum information is axiomatic. Its richness and elegance is demonstrated in that only four axioms make it complete.¹

I.1 Quantum information axioms

I.1.1 The superposition axiom

A quantum unit (qubit) is specified by a two element vector, say $\begin{bmatrix} \alpha \\ \beta \end{bmatrix}$, with $|\alpha|^2 + |\beta|^2 = 1$.

¹Some argue that the measurement axiom is contained in the transformation axiom, hence requiring only three axioms. Even though we appreciate the merits of the argument, we'll proceed with four axioms. Feel free to count them as only three if you prefer.

Let $|\psi\rangle \equiv \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \alpha|0\rangle + \beta|1\rangle$,² $\langle\psi| = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}^\dagger$ where \dagger is the adjoint (conjugate transpose) operation.

I.1.2 The transformation axiom

A transformation of a quantum unit is accomplished by unitary (length-preserving) matrix multiplication. The Pauli matrices provide a basis of unitary operators.

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

$$Y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \quad Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

where $i = \sqrt{-1}$. The operations work as follows: $I \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$, $X \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} \beta \\ \alpha \end{bmatrix}$, $Y \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = -\begin{bmatrix} \beta i \\ \alpha i \end{bmatrix}$, and $Z \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = -\begin{bmatrix} \alpha \\ \beta \end{bmatrix}$. Other useful single qubit transformations are $H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ and $\Theta = \begin{bmatrix} e^{i\theta} & 0 \\ 0 & 1 \end{bmatrix}$. Examples of these transformations in Dirac notation are³

$$H|0\rangle = \frac{|0\rangle + |1\rangle}{\sqrt{2}}; \quad H|1\rangle = \frac{|0\rangle - |1\rangle}{\sqrt{2}}$$

$$\Theta|0\rangle = e^{i\theta}|0\rangle; \quad \Theta|1\rangle = |1\rangle$$

I.1.3 The measurement axiom

Measurement occurs via interaction with the quantum state as if a linear projection is applied to the quantum state.⁴ The set of projection matrices

²Dirac notation is a useful descriptor, as $|0\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $|1\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$.

³A summary table for common quantum operators expressed in Dirac notation is provided in section I.2.

⁴This is a compact way of describing measurement. However, conservation of information (a principle of quantum information) demands that operations be reversible. In other words, all transformations (including interactions to measure) be unitary — projections are not unitary. However, there always exist unitary operators that produce the same post-measurement state as that indicated via projection. Hence, we treat projections as an expedient for describing measurement.

are complete as they add to the identity matrix.

$$\sum_m M_m^\dagger M_m = I$$

where M_m^\dagger is the adjoint (conjugate transpose) of projection matrix M_m . The probability of a particular measurement occurring is the squared absolute value of the projection. (An implication of the axiom not explicitly used here is that the post-measurement state is the projection appropriately normalized; this effectively rules out multiple measurement.)

For example, let the projection matrices be $M_0 = |0\rangle\langle 0| = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ and $M_1 = |1\rangle\langle 1| = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$. Note that M_0 projects onto the $|0\rangle$ vector and M_1 projects onto the $|1\rangle$ vector. Also note that $M_0^\dagger M_0 + M_1^\dagger M_1 = M_0 + M_1 = I$. For $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, the projection of $|\psi\rangle$ onto $|0\rangle$ is $M_0|\psi\rangle$. The probability of $|0\rangle$ being the result of the measurement is $\langle\psi|M_0|\psi\rangle = |\alpha|^2$.

I.1.4 The combination axiom

Qubits are combined by tensor multiplication. For example, two $|0\rangle$ qubits

are combined as $|0\rangle \otimes |0\rangle = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$ denoted $|00\rangle$. It is often useful to trans-

form one qubit in a combination and leave another unchanged; this can also be accomplished by tensor multiplication. Let H_1 denote a Hadamard transformation on the first qubit. Then applied to a two qubit system,

$$H_1 = H \otimes I = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix} \text{ and } H_1|00\rangle = \frac{|00\rangle + |10\rangle}{\sqrt{2}}.$$

Another important two qubit transformation is the controlled not operator,

$$Cnot = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

The controlled not operator flips the target, second qubit, if the control, first qubit, equals $|1\rangle$ and otherwise leaves the target unchanged: $Cnot|00\rangle = |00\rangle$, $Cnot|01\rangle = |01\rangle$, $Cnot|10\rangle = |11\rangle$, and $Cnot|11\rangle = |10\rangle$,

Entangled two qubit states or Bell states are defined as follows,

$$|\beta_{00}\rangle = Cnot H_1 |00\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}}$$

and more generally,

$$|\beta_{ij}\rangle = Cnot H_1 |ij\rangle \text{ for } i, j = 0, 1$$

The four two qubit Bell states form an orthonormal basis.

I.2 Summary of quantum "rules"

Below we tabulate two qubit quantum operator rules. The column heading indicates the initial state while the row value corresponding the operator identifies the transformed state. Of course, the same rules apply to one qubit (except *Cnot*) or many qubits, we simply have to exercise care to identify which qubit is transformed by the operator (we continue to denote the target qubit via the subscript on the operator).

operator	$ 00\rangle$	$ 01\rangle$	$ 10\rangle$	$ 11\rangle$
I_1 or I_2	$ 00\rangle$	$ 01\rangle$	$ 10\rangle$	$ 11\rangle$
X_1	$ 10\rangle$	$ 11\rangle$	$ 00\rangle$	$ 01\rangle$
X_2	$ 01\rangle$	$ 00\rangle$	$ 11\rangle$	$ 10\rangle$
Z_1	$ 00\rangle$	$ 01\rangle$	$- 10\rangle$	$- 11\rangle$
Z_2	$ 00\rangle$	$- 01\rangle$	$ 10\rangle$	$- 11\rangle$
Y_1	$i 10\rangle$	$i 11\rangle$	$-i 00\rangle$	$-i 01\rangle$
Y_2	$i 01\rangle$	$-i 00\rangle$	$i 11\rangle$	$-i 10\rangle$
H_1	$\frac{ 00\rangle+ 10\rangle}{\sqrt{2}}$	$\frac{ 01\rangle+ 11\rangle}{\sqrt{2}}$	$\frac{ 00\rangle- 10\rangle}{\sqrt{2}}$	$\frac{ 01\rangle- 11\rangle}{\sqrt{2}}$
H_2	$\frac{ 00\rangle+ 01\rangle}{\sqrt{2}}$	$\frac{ 00\rangle- 01\rangle}{\sqrt{2}}$	$\frac{ 10\rangle+ 11\rangle}{\sqrt{2}}$	$\frac{ 10\rangle- 11\rangle}{\sqrt{2}}$
Θ_1	$e^{i\theta} 00\rangle$	$e^{i\theta} 01\rangle$	$ 10\rangle$	$ 11\rangle$
Θ_2	$e^{i\theta} 00\rangle$	$ 01\rangle$	$e^{i\theta} 10\rangle$	$ 11\rangle$
<i>Cnot</i>	$ 00\rangle$	$ 01\rangle$	$ 11\rangle$	$ 10\rangle$

common quantum operator rules

I.3 Observables and expected payoffs

Measurements involve real values drawn from observables. To ensure measurements lead to real values, observables are also represented by Hermitian matrices. A Hermitian matrix is one in which the complex conjugate of the matrix equals the original matrix, $M^\dagger = M$. Hermitian matrices have real eigenvalues and eigenvalues are the values realized via measurement. Suppose we are working with observable M in state $|\psi\rangle$ where

$$\begin{aligned} M &= \begin{bmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \lambda_3 & 0 \\ 0 & 0 & 0 & \lambda_4 \end{bmatrix} \\ &= \lambda_1 |00\rangle\langle 00| + \lambda_2 |01\rangle\langle 01| + \lambda_3 |10\rangle\langle 10| + \lambda_4 |11\rangle\langle 11| \end{aligned}$$

The expected payoff is

$$\begin{aligned} \langle M \rangle &= \langle \psi | M | \psi \rangle \\ &= \lambda_1 \langle \psi | 00 \rangle \langle 00 | \psi \rangle + \lambda_2 \langle \psi | 01 \rangle \langle 01 | \psi \rangle \\ &\quad + \lambda_3 \langle \psi | 10 \rangle \langle 10 | \psi \rangle + \lambda_4 \langle \psi | 11 \rangle \langle 11 | \psi \rangle \end{aligned}$$

In other words, λ_1 is observed with probability $\langle \psi | 00 \rangle \langle 00 | \psi \rangle$, λ_2 is observed with probability $\langle \psi | 01 \rangle \langle 01 | \psi \rangle$, λ_3 is observed with probability $\langle \psi | 10 \rangle \langle 10 | \psi \rangle$, and λ_4 is observed with probability $\langle \psi | 11 \rangle \langle 11 | \psi \rangle$.

I.4 Density operators and quantum entropy

To this point we've focused on pure states. How do we proceed if our state of knowledge indicates a mixture of states (that is, a probability weighted average of states)? Density operators supply the frame for mixed as well as pure states. A density operator is defined as

$$\rho = \sum_{i=1}^n p_i |\psi_i\rangle \langle \psi_i|$$

where the trace (sum of the diagonal elements) equals one, $tr(\rho) = 1$. For density operator ρ the expected value associated with observable $M = \sum_{k=1}^n \lambda_k |k\rangle \langle k|$ (via spectral decomposition where $|k\rangle$ is an orthonormal basis) is $\langle M \rangle = tr(\rho M) = tr(M\rho)$.

This follows as the probability of observing λ_k equals

$$\Pr(\lambda_k) = \langle k | \rho | k \rangle = \sum_{i=1}^n p_i \langle k | \psi_i \rangle \langle \psi_i | k \rangle$$

and useful properties of the trace.

$$tr(BA) = tr(AB)$$

To see this, recognize $(AB)_{ik} = \sum_j a_{ij} b_{jk}$ then $tr(AB) = \sum_{i,j} a_{ij} b_{ji}$ and $tr(BA) = \sum_{i,j} b_{ji} a_{ij}$ which is the same as $\sum_{i,j} a_{ij} b_{ji}$. Now, let $B \equiv |\psi\rangle \langle \psi|$

$$\begin{aligned} tr(AB) &= tr(A|\psi\rangle \langle \psi|) \\ &= \sum_i \langle i | A | \psi \rangle \langle \psi | i \rangle \\ &= \langle \psi | A | \psi \rangle \end{aligned}$$

where $|i\rangle$ is an orthonormal basis for $|\psi\rangle$ with first element $|\psi\rangle$.⁵ The second line implements $tr(AB) = \sum_{i,j} a_{ij} b_{ji}$ while the third line follows from orthogonality of the basis for $|\psi\rangle$, that is, $\langle \psi | i \rangle$ is either 0 or 1.

⁵Notice for a pure state $|\psi\rangle$, $\langle \psi | A | \psi \rangle = \langle A \rangle_{|\psi\rangle}$, the expected value of the observable when the system is in state $|\psi\rangle$ equals $tr(A|\psi\rangle \langle \psi|)$.

Then, applying the first result followed by the third result in reverse and repeating based on the second result we have

$$\begin{aligned}
\langle M \rangle &= \sum_{k=1}^n \text{Pr}(\lambda_k) \lambda_k \\
&= \sum_{k=1}^n \sum_{i=1}^n p_i \langle k | \psi_i \rangle \langle \psi_i | k \rangle \lambda_k \\
&= \sum_{k=1}^n \langle k | \rho | k \rangle \lambda_k \\
&= \text{tr}(M\rho) \\
&= \sum_{k=1}^n \sum_{i=1}^n p_i \langle \psi_i | k \rangle \langle k | \psi_i \rangle \lambda_k \\
&= \sum_{i=1}^n p_i \langle \psi_i | M | \psi_i \rangle \\
&= \text{tr}(\rho M)
\end{aligned}$$

Consider an example. Suppose

$$\rho = \begin{bmatrix} 0.4 & 0 \\ 0 & 0.6 \end{bmatrix} = 0.4 \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \end{bmatrix} + 0.6 \begin{bmatrix} 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \end{bmatrix}$$

and

$$M = \begin{bmatrix} 1 & -2 \\ -2 & 1 \end{bmatrix} = 3 \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} - 1 \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

Then,

$$\begin{aligned}
\text{Pr}(\lambda = 3) &= \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 0.4 & 0 \\ 0 & 0.6 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix} \\
&= 0.4 \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix} \\
&\quad + 0.6 \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix} \\
&= \frac{1}{2}
\end{aligned}$$

and

$$\begin{aligned}
 \Pr(\lambda = -1) &= \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 0.4 & 0 \\ 0 & 0.6 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \\
 &= 0.4 \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \\
 &\quad + 0.6 \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \\
 &= \frac{1}{2}
 \end{aligned}$$

Hence, the expected value for the observable M is

$$\begin{aligned}
 \langle M \rangle &= \sum_{k=1}^2 \Pr(\lambda_k) \lambda_k \\
 &= \frac{1}{2} 3 + \frac{1}{2} (-1) = 1 \\
 &= \text{tr}(\rho M) = \text{tr}(M \rho) \\
 &= \begin{bmatrix} 0.4 & -0.8 \\ -1.2 & 0.6 \end{bmatrix} = \begin{bmatrix} 0.4 & -1.2 \\ -0.8 & 0.6 \end{bmatrix} = 1
 \end{aligned}$$

or

$$\begin{aligned}
 \langle M \rangle &= \sum_{i,j} p_j \langle i | M | \psi \rangle \langle \psi | i \rangle \\
 &= 0.4 \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & -2 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\
 &\quad + 0.4 \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -2 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\
 &\quad + 0.6 \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & -2 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\
 &\quad + 0.6 \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -2 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\
 &= \sum_j p_j \langle \psi | M | \psi \rangle \\
 &= 0.4 \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & -2 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\
 &\quad + 0.6 \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -2 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\
 &= 0.4 + 0.6 = 1
 \end{aligned}$$

I.4.1 Quantum entropy

von Neumann defines quantum entropy as

$$S = -\text{tr}(\rho \log \rho)$$

For pure states, ρ is a Hermitian matrix whose logarithm (discussed earlier) involves its spectral decomposition $\rho = Q\Lambda Q^T$ where Λ has one eigenvalue equal to unity and the remainder equal to zero. Hence, $\log \rho = Q \log \Lambda Q^T = 0$ (by convention, $0 \log 0 = 0$) so that von Neumann entropy is zero for pure states.

On the other hand, for mixed states spectral decomposition of ρ is $\sum_{j=1}^n \lambda_j |j\rangle \langle j|$. This involves nonzero eigenvalues so that

$$\begin{aligned} S &= -\text{tr}(\rho \log \rho) \\ &= -\text{tr} \left(\sum_{j=1}^n \lambda_j |j\rangle \langle j| \sum_{j=1}^n \log(\lambda_j) |j\rangle \langle j| \right) \\ &= -\text{tr} \left(\sum_{j=1}^n \lambda_j \log(\lambda_j) |j\rangle \langle j| |j\rangle \langle j| \right) \\ &= -\text{tr} \left(\sum_{j=1}^n \lambda_j \log(\lambda_j) |j\rangle \langle j| \right) \\ &= -\sum_{i=1}^n \lambda_i \log \lambda_i \end{aligned}$$

The latter result follows from equality of the sum of the eigenvalues and the trace of a matrix.

I.5 Some trigonometric identities

Most trigonometric identities follow directly from Euler's equation:

$$e^{\pm i\theta} = \cos \theta \pm i \sin \theta$$

Remark 1 $\cos^2 \theta + \sin^2 \theta = 1$

Proof.

$$\begin{aligned} e^{i\theta} e^{-i\theta} &= e^0 = 1 \\ (\cos \theta + i \sin \theta)(\cos \theta - i \sin \theta) &= \cos^2 \theta + i \cos \theta \sin \theta - i \cos \theta \sin \theta + \sin^2 \theta \\ &= \cos^2 \theta + \sin^2 \theta \end{aligned}$$

■

Remark 2 $\frac{1-\cos\theta}{2} = \sin^2 \frac{\theta}{2}$

Proof.

$$\begin{aligned} \frac{2 - (e^{i\theta} + e^{-i\theta})}{4} &= \frac{2 - 2\cos\theta}{4} = \frac{1 - \cos\theta}{2} \\ \frac{2 - \left[\left(e^{\frac{i\theta}{2}} \right)^2 + \left(e^{-\frac{i\theta}{2}} \right)^2 \right]}{4} &= \frac{2 - [2 - 4\sin^2 \frac{\theta}{2}]}{4} \\ &= \sin^2 \frac{\theta}{2} \end{aligned}$$

Details related to the second line are below.

$$\begin{aligned} \left(e^{\frac{i\theta}{2}} \right)^2 &= \left(\cos \frac{\theta}{2} + i \sin \frac{\theta}{2} \right)^2 \\ &= \cos^2 \frac{\theta}{2} - \sin^2 \frac{\theta}{2} + 2i \cos \frac{\theta}{2} \sin \frac{\theta}{2} \end{aligned}$$

$$\begin{aligned} \left(e^{-\frac{i\theta}{2}} \right)^2 &= \left(\cos \frac{\theta}{2} - i \sin \frac{\theta}{2} \right)^2 \\ &= \cos^2 \frac{\theta}{2} - \sin^2 \frac{\theta}{2} - 2i \cos \frac{\theta}{2} \sin \frac{\theta}{2} \end{aligned}$$

$$\begin{aligned} \left(e^{\frac{i\theta}{2}} \right)^2 + \left(e^{-\frac{i\theta}{2}} \right)^2 &= 2\cos^2 \frac{\theta}{2} - 2\sin^2 \frac{\theta}{2} \\ &= 2 - 4\sin^2 \frac{\theta}{2} \end{aligned}$$

Remark 3 $\frac{1-\cos\theta_1\cos\theta_2}{2} = \cos^2 \frac{\theta_1}{2} \sin^2 \frac{\theta_2}{2} + \sin^2 \frac{\theta_1}{2} \cos^2 \frac{\theta_2}{2}$

■
Proof. From the two preceding identities

$$\frac{1 - \cos\theta}{2} = \sin^2 \frac{\theta}{2} = 1 - \cos^2 \frac{\theta}{2}$$

and

$$\begin{aligned} \cos\theta &= 1 - 2\sin^2 \frac{\theta}{2} \\ &= 2\cos^2 \frac{\theta}{2} - 1 \end{aligned}$$

Therefore,

$$\begin{aligned}
\frac{1 - \cos \theta_1 \cos \theta_2}{2} &= \frac{1 - (1 - 2 \sin^2 \frac{\theta_1}{2}) (2 \cos^2 \frac{\theta_2}{2} - 1)}{2} \\
&= \frac{1 - (2 \cos^2 \frac{\theta_2}{2} - 2 \sin^2 \frac{\theta_1}{2} 2 \cos^2 \frac{\theta_2}{2} - 1 + 2 \sin^2 \frac{\theta_1}{2})}{2} \\
&= -\cos^2 \frac{\theta_2}{2} + 2 \sin^2 \frac{\theta_1}{2} \cos^2 \frac{\theta_2}{2} + 1 - \sin^2 \frac{\theta_1}{2} \\
&= \cos^2 \frac{\theta_2}{2} \left(2 \sin^2 \frac{\theta_1}{2} - 1 \right) + \cos^2 \frac{\theta_1}{2} \\
&= \cos^2 \frac{\theta_2}{2} \left(2 \sin^2 \frac{\theta_1}{2} - \cos^2 \frac{\theta_1}{2} - \sin^2 \frac{\theta_1}{2} \right) + \cos^2 \frac{\theta_1}{2} \\
&= \cos^2 \frac{\theta_2}{2} \left(\sin^2 \frac{\theta_1}{2} - \cos^2 \frac{\theta_1}{2} \right) + \cos^2 \frac{\theta_1}{2} \\
&= \cos^2 \frac{\theta_2}{2} \sin^2 \frac{\theta_1}{2} - \cos^2 \frac{\theta_2}{2} \cos^2 \frac{\theta_1}{2} + \cos^2 \frac{\theta_1}{2} \\
&= \sin^2 \frac{\theta_1}{2} \cos^2 \frac{\theta_2}{2} + \cos^2 \frac{\theta_1}{2} \left(1 - \cos^2 \frac{\theta_2}{2} \right) \\
&= \cos^2 \frac{\theta_1}{2} \sin^2 \frac{\theta_2}{2} + \sin^2 \frac{\theta_1}{2} \cos^2 \frac{\theta_2}{2}
\end{aligned}$$

■

Remark 4 $\sin 2\theta = 2 \cos \theta \sin \theta$

Proof.

$$\begin{aligned}
\frac{e^{i2\theta} - e^{-i2\theta}}{2i} &= \frac{\cos 2\theta + i \sin 2\theta - (\cos 2\theta - i \sin 2\theta)}{2i} \\
&= \frac{2i \sin 2\theta}{2i} \\
&= \sin 2\theta \\
\frac{(e^{i\theta})^2 - (e^{-i\theta})^2}{2i} &= \frac{(\cos \theta + i \sin \theta)^2 - (\cos \theta - i \sin \theta)^2}{2i} \\
&= \frac{\cos^2 \theta + 2i \cos \theta \sin \theta - \sin^2 \theta - (\cos^2 \theta - 2i \cos \theta \sin \theta - \sin^2 \theta)}{2i} \\
&= \frac{4i \cos \theta \sin \theta}{2i} \\
&= 2 \cos \theta \sin \theta
\end{aligned}$$

■