1. Definition: Plus Minus Basis	2. Definition: Operator Norm
3. Definition: Measurement	4. Definition: Measurement.Projective
5. Definition: Observable	6. Definition: Operator.Positive Semi Definite
7. Definition: Povm	8. Definition: Pauli Matrices
9. Definition: Trace	10. Proposition: Expectation As Trace
11. Definition: Density Matrix	12. Postulate: Heisenberg: State Space
13. Postulate: Heisenberg: Evolution	14. Definition: Hamiltonian Ground State

15. Postulate: Heisenberg: Measurement	16. Postulate: Heisenberg: Composite System
17. Definition: Entangled State	18. Postulate: Schrodinger: System State
19. Postulate: Schrodinger: Evolution	20. Postulate: Schrodinger: Measurement
21. Postulate: Schrodinger: Composite System	22. Theorem: Schmidt Decomposition
23. Proposition: Schmidt Rank Entanglement Criterion	24. Definition: State.Maximally Entangled
25. Definition: Partial Trace	26. Definition: Reduced Density Matrix

27. Proposition: Equivalent Definition Of Partial Trace	28. Proposition: Partial Trace Properties
29. Definition: Purification	30. Theorem: Naimark
31. Definition: Quantum Channel	32. Definition: Maximally Entangled State
33. Definition: Partial Transpose	34. Definition: Hilbert Schmidt Inner Product
35. Definition: Superoperator Adjoint	36. Definition: Choi Jamiolkowski Matrix
37. Theorem: Characterisation Of Quantum Channels	38. Definition: Depolarising Channel

39. Definition: Phase Damping Channel	40. Definition: Separable Density Matrix
41. Definition: Entanglement Breaking	42. Proposition: Equivalence Of Ensembles
43. Theorem: Radon Nikodym	44. Definition: Instrument
45. Proposition: Quantum Steering	46. Definition: Quantum Markov Semigroup
47. Definition: Markovian Quantum Master Equation	48. Definition: Operator Correlation
49. Definition: Entanglement Entropy	50. Proposition: Properties Of Entanglement Entropy

51. Proposition: Ppt Criterion	52. Definition: Entanglement Witness
53. Proposition: Positive Map Criterion For Separability	54. Definition: Decomposable Map
55. Proposition: Reduction Criterion	56. Proposition: Entangled States With Ppt Exist Iff Non Decomposable Maps Exist
57. Proposition: Separability Criterion In Small Dimensions	58. Definition: Optimal Measurement
59. Proposition: Mle Always Exists	60. Theorem: Equivalent Conditions For Optimal Measurements

61. Definition: Schatten P Norm	62. Theorem: Quantum Neyman Pearson
63. Lemma: Quantum Chernoff Bound Lemma	64. Theorem: Quantum Chernoff Bound
65. Definition: Pretty Good Measurement	66. Definition: Square Measurement
67. Theorem: Holders Inequality	68. Definition: Operator Convex
69. Theorem: Jensens Inequality	70. Proposition: Square Measurement Probability Bounds
71. Proposition: Pretty Good Measurement Probability Bounds	72. Corollary: Pretty Good And Square Success And Error Bounds

73. Definition: Type I And Ii Errors	74. Theorem: Quantum Steins Lemma
75. Definition: Von Neumann Entropy	76. Proposition: Properties Of Von Neumann Entropy
77. Corollary: Second Law Of Thermodynamics	78. Definition: Quantum Mutual Information
79. Definition: Conditional Mutual Information	80. Definition: Conditional Entropy
81. Proposition: Bounds On Conditional Entropy	82. Definition: Umegaki Relative Entropy
83. Definition: Belavkin Staszewski Relative Entropy	84. Proposition: Umegaki Entropy Upper Bound

85. Proposition: Properties Of Quantum Relative Entropy	86. Definition: RéNyi Divergence
87. Definition: Axioms Of Quantum Renyi Divergence	88. Definition: Pinching Map
89. Definition: Preparation Map	90. Definition: Minimal RéNyi Divergence
91. Definition: Maximal Renyi Divergence	92. Proposition: Properties Of Maximal And Minimal Renyi Divergence
93. Definition: Petz RéNyi Divergence	94. Definition: Max Relative Entropy

95. Definition: Quantum Divergences Ordering	96. Definition: Quantum Channel Divergence
97. Lemma: Matsumoto	98. Theorem: Chain Rule For Quantum Channels
99. Definition: Regularised Renyi Divergence	100. Theorem: Schumacher

1. Basic notions in quantum information theory

The field is motivated by the fact that we want to control quantum systems.

1. Can we construct and manipulate quantum systems?
2. If so, which are the scientific and technological applications?

Entanglement frontier: highly complex quantum systems, which are more complex and richer than classical systems. However, quantum systems have *decoherence*, which classical systems don't. "Quantum advantage" gives speed up over classical systems.

Quantum vs classical information theory:

- True randomness.

- Uncertainty.
- Entanglement.

Note we always work with finite-dimensional Hilbert spaces, so take $\mathbb{H} = \mathbb{C}^N$.

1.1. Qubits and basic operations

Notation 1.1 Vectors are denoted by $|\psi\rangle \in \mathbb{C}^n$, dual vectors by $\langle\psi| \in (\mathbb{C}^n)^*$, and inner products by $\langle\psi|\phi\rangle \in \mathbb{C}$. $|\psi\rangle\langle\psi| : \mathbb{C}^n \to \mathbb{C}^n$ are rank-one projectors.

Definition: Plus Minus Basis

Definition 1.2 Another important basis of $\mathbb{C}^2$ is $\{|+\rangle, |-\rangle\}$, where $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ and $|-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$.

Definition: Operator Norm

Definition 1.3 For an operator $T : \mathbb{H} \to \mathbb{H}$, the **operator norm** of T is

$$\|T\| = \|T\|_{\mathbb{H}\to\mathbb{H}} := \sup_{x\in H} \frac{\|T(x)\|_{\mathbb{H}}}{\|x\|_{\mathbb{H}}}$$

Notation 1.4 Let $\mathbb{B}(\mathbb{H})$ denote the space of bounded linear operators, i.e. T such that $\|T\| < \infty$.

Notation 1.5 Denote the dual of the operator T by T^*, i.e. the operator that satisfies $\langle y|T(x)\rangle = \langle T^*(y)|x\rangle$ for all $x, y \in \mathbb{H}$.

Definition: Measurement

Definition 1.6 A **quantum measurement** is a collection of measurement operators $\{M_n\}_n \subseteq \mathbb{B}(\mathbb{H})$ which satisfies $\sum_n M_n^* M_n = \mathbb{I}$, the identity operator.

Given $|\phi\rangle$, the probability that $|n\rangle$ occurs after this operation is $p(n) = \langle\phi|M_n^* M_n|\phi\rangle$. After performing this operation, the state of the system is $\frac{1}{\sqrt{p(n)}} M_n|\phi\rangle$. This is the **Born rule**.

Example 1.7 A measurement in the computational basis is $M_0 = |0\rangle\langle 0|$, $M_1 = |1\rangle\langle 1|$. Note M_0 and M_1 are self-adjoint. Let $|\psi\rangle = \alpha_0|0\rangle + \alpha_1|1\rangle$. Then $p(i) = \langle\phi|M_i|\phi\rangle = |\alpha_i|^2$. The state after measurement is $\frac{\alpha_i}{|\alpha_i|}|i\rangle$, which is equivalent to $|i\rangle$.

Note that $|\psi\rangle$ and $e^{i\theta}|\psi\rangle$ are operationally identical: the phase does not affect the measurement probabilities.

Definition: Measurement.Projective

Definition 1.8 A quantum measurement $\{M_n\}_n \subseteq \mathbb{B}(\mathbb{H})$ is **projective measurement** if the M_n are orthogonal projections (i.e. they are self-adjoint (Hermitian) and $M_n M_m = \delta_{nm} M_n$).

Definition: Observable

Definition 1.9 An **observable** is a Hermitian operator, which we can express as its spectral decomposition

$$M = \sum_n \lambda_n M_n,$$

where $\{M_n\}_n$ is a projective measurement. The possible outcomes of the measurement correspond to its eigenvalues λ_n of the observable. Note that the expected value of the measurement is

$$\sum_n \lambda_n p(n) = \sum_n \lambda_n \langle \phi | M_n | \phi \rangle = \langle \phi | M | \phi \rangle.$$

Definition: Operator.Positive Semi Definite

Definition 1.10 $T : \mathbb{H} \to \mathbb{H}$ is **positive (semi-definite)** (written $T \geq 0$) if $\langle\psi|T|\psi\rangle \geq 0$ for all $|\psi\rangle \in H$.

Definition: Povm

Definition 1.11 A **POVM (positive operator valued measurement)** is a collection $\{E_n\}_n$ where each $E_n = M_n^* M_n$ for a general measurement $\{M_n\}_n$ (i.e. each E_n is positive and Hermitian, and $\sum_n E_n = \mathbb{I}$).

Note that the probability of obtaining outcome m on $|\psi\rangle$ is $p(m) = \langle\psi|E_m|\psi\rangle$. We use POVMs when we care only about the probabilities of the different measurement outcomes, and not the post-measurement states.

Conversely, given a POVM $\{E_n\}_n$, we can define a general measurement $\{\sqrt{E_n}\}_n$.

Remark 1.12 Any transformation on a normalised quantum state must map it to a normalised quantum state, and so the operation must be unitary.

Definition: Pauli Matrices

Definition 1.13 The **Pauli matrice** are

$$\sigma_0 = \mathbb{I} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \sigma_X = X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

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

The Pauli matrices are unitaries, and we can think of them as quantum logical gates.

Definition: Trace

Definition 1.14 The **trace** of $T : \mathbb{H} \to \mathbb{H}$ is

$$\operatorname{tr} T = \operatorname{tr} M = \sum_i M_{ii} \in \mathbb{C},$$

where M is a matrix representation of T in any basis (this is well-defined since the trace is cyclic and linear).

Proposition: Expectation As Trace

Proposition 1.15 For any state $|\phi\rangle$ and any operator A,

$$\mathrm{tr}(A|\phi\rangle\langle\phi|) = \langle\phi|A|\phi\rangle.$$

Proof (Hints). Straightforward. □

Proof. $\mathrm{tr}(A|\phi\rangle\langle\phi|) = \sum_i \langle i|A|\phi\rangle\langle\phi|i\rangle$ for an orthonormal basis $\{|i\rangle\}$. Any basis where $|\phi\rangle = |j\rangle$ for some j instantly yields the result. Alternatively, we have

$$\mathrm{tr}(A|\phi\rangle\langle\phi|) = \sum_i \langle i|A|\phi\rangle\langle\phi|i\rangle = \sum_i \langle\phi|i\rangle\langle i|A|\phi\rangle = \langle\phi|I|A|\phi\rangle = \langle\phi|A|\phi\rangle.$$

□

Suppose we don't fully know the state of the system, but know that it is $|\phi_i\rangle$ with probability p_i. We want to be able to consider the $\sum_i p_i|\phi_i\rangle$ as a state, but this isn't normalised (except when some $p_i = 1$). To solve this issue, we assume each $|\phi_i\rangle$ to the rank-one projector $|\phi_i\rangle\langle\phi_i|$, and we describe the unknown state by $\rho = \sum_i p_i|\phi_i\rangle\langle\phi_i|$. This gives rise to the following definition:

Definition: Density Matrix

Definition 1.16 A **density matrix/operator** is a linear operator $\rho \in \mathbb{B}(\mathbb{H})$ which is:

- Hermitian,
- Positive semi-definite, and
- Satisfies $\operatorname{tr} \rho = 1$.

1.2. Postulates of quantum mechanics (Heisenberg picture)

Postulate: Heisenberg: State Space

Postulate 1.17 Given an isolated physical system, there exists a complex (separable) Hilbert space $\mathbb{H}$ associated with it, called **state space**. The physical system is described by a **state vector**, which is a normalised vector in $\mathbb{H}$.

Postulate: Heisenberg: Evolution

Postulate 1.18 Given an isolated physical system, its evolution is described by a unitary. If the state of the system at time t_1 is $|\phi_1\rangle$ and at time t_2 is $|\phi_2\rangle$, then there exists a unitary U_{t_1,t_2} such that $|\phi_2\rangle = U_{t_1,t_2}|\phi_1\rangle$.

This can be generalised with the Schrodinger equation: the time evolution of a closed quantum system is given by $i\hbar\frac{\mathrm{d}}{\mathrm{d}t}|\phi(t)\rangle = H|\phi(t)\rangle$. The Hermitian operator H is called the **Hamiltonian** and is generally time-dependent.

Definition: Hamiltonian Ground State

Definition 1.19 Let the spectral decomposition of H be

$$H = \sum_i E_i |E_i\rangle\langle E_i|,$$

where the E_i are the **energy eigenvalues** and the $|E_i\rangle$ are the **energy eigenstates** (or **stationary states**).

The minimum energy is called the **ground state energy** and its associated eigenstate is called the **ground state**. The **(spectral) gap** of H is the (absolute) difference between the ground state energy and the next largest energy eigenvalue. When the gap is strictly positive,

we say the system is **gapped**. The states $|E_i\rangle$ are called **stationary**, since they evolve as $|E_i\rangle \to \exp(-iE_i t/\hbar)|E_i\rangle$.

We have $|\phi(t_2)\rangle = U(t_1, t_2)|\phi(t_1)\rangle$ where $U(t_1, t_2) = \exp(-iH(t_2 - t_1)/\hbar)$ which is a unitary. In fact, any unitary U can be written in the form $U = \exp(iK)$ for some Hermitian K.

Postulate: Heisenberg: Measurement

Postulate 1.20 Given a physical system with associated Hilbert space $\mathbb{H}$, quantum measurements in the system are described by a collection of measurements $\{M_n\}_n \subseteq \mathbb{B}(\mathbb{H})$ such that $\sum_n M_n^* M_n = \mathbb{I}$, as in Definition 1.6. The index n refers to the measurement outcomes that may occur in the experiment, and given a state $|\phi\rangle$ before measurement, the probability that n occurs is

$$p(n) = \langle\phi|M_n^* M_n|\phi\rangle.$$

The state of the system after measurement is $\frac{1}{\sqrt{p(n)}} M_n|\phi\rangle$

Postulate: Heisenberg: Composite System

Postulate 1.21 Given a composite physical system, its state space $\mathbb{H}$ is also composite and corresponds to the tensor product of the individual state spaces $\mathbb{H}_i$ of each component: $\mathbb{H} = \mathbb{H}_1 \otimes \cdots \otimes \mathbb{H}_N$. If the state in each system i is $|\phi_i\rangle$, then the state in the composite system is $|\phi_1\rangle \otimes \cdots \otimes |\phi_N\rangle$.

Definition: Entangled State

Definition 1.22 Given $|\phi\rangle \in H_1 \otimes \cdots \otimes H_N$, $|\phi\rangle$ is **entangled** if it cannot be written as a tensor product of the form $|\phi_1\rangle \otimes \cdots \otimes |\phi_n\rangle$. Otherwise, it is **separable** or a **product state**.

Example 1.23 The **EPR pair (Bell state)** $|\phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ is entangled.

1.3. Postulates of quantum mechanics (Schrodinger picture)

Postulate: Schrodinger: System State

Postulate 1.24 Given an isolated physical system, the state of the system is completely described by its density operator, which is Hermitian, positive semi-definite and has trace one.

If we know the system is in state ρ_i with probability p_i, then the state of the system is $\sum_i p_i \rho_i$.

Pure states are of the form $\rho = |\phi\rangle\langle\phi|$, **mixed states** are of the form $\rho = \sum_i p_i |\phi_i\rangle\langle\phi_i|$.

Postulate: Schrodinger: Evolution

Postulate 1.25 Given an isolated physical system, its evolution is described by a unitary. If the state of the system is ρ_1 at time t_1 and is ρ_2 at time t_2, then there is a unitary U depending only on t_1, t_2 such that $\rho_2 = U\rho_1 U^*$.

Postulate: Schrodinger: Measurement

Postulate 1.26 The same as Postulate 1.20, except we specify that after measurement $\{M_n\}_n$, the probability of observing n is $p(n) = \mathrm{tr}(M_n^* M_n \rho)$ and the state after measurement is $\frac{1}{p(n)} M_n \rho M_n^*$.

Postulate: Schrodinger: Composite System

Postulate 1.27 The same as Postulate 1.21, except that the state of the composite system is $\rho = \rho_1 \otimes \cdots \otimes \rho_n$, where ρ_i is the state of ith individual system.

Remark 1.28 The Heisenberg and Schrodinger postulates are mathematically equivalent.

1.4. States, entanglement and measurements

Theorem: Schmidt Decomposition

Theorem 1.29 (Schmidt Decomposition) Let $|\psi\rangle$ be a pure state in a bipartite system $\mathbb{H}_{AB} = \mathbb{H}_A \otimes \mathbb{H}_B$, where $\mathbb{H}_A$ has dimension N_A and $\mathbb{H}_B$ has dimension $N_B \geq N_A$. Then there exist orthonormal states $\{|e_i\rangle : i \in [N_A]\} \subseteq \mathbb{H}_A$ and $\{|f_i\rangle : i \in [N_A]\} \subseteq \mathbb{H}_B$ such that

$$|\psi\rangle = \sum_{i=1}^{N_A} \lambda_i |e_i\rangle \otimes |f_i\rangle,$$

where $\lambda_i \geq 0$ and $\sum_i \lambda_i^2 = 1$.

The λ_i are unique up to re-ordering. The λ_i are called the **Schmidt coefficients** and the number of $\lambda_i > 0$ is the **Schmidt rank** (or **Schmidt number**) of the state.

Proof (Hints). Use the singular value decomposition of the matrix of amplitudes of $|\psi\rangle$. $\square$

Proof. Let $|\psi\rangle = \sum_{k=1}^{N_A} \sum_{\ell=1}^{N_B} \beta_{k\ell} |\phi_k\rangle \otimes |\phi_\ell\rangle$ for orthonormal bases $\{|\phi_k\rangle : k \in [N_A]\} \subseteq \mathbb{H}_A$, $\{|\chi_\ell\rangle : \ell \in [N_B]\} \subseteq \mathbb{H}_B$. Let $(\beta_{k\ell})$ have singular value decomposition

$$U[\Sigma \ 0]V,$$

where U is an $N_A \times N_A$ unitary, Σ is an $N_A \times N_A$ diagonal matrix with non-negative entries, and V is an $N_B \times N_B$ unitary. So

$$\beta_{k\ell} = \sum_{i=1}^{N_A} \sum_{j=1}^{N_B} U_{ki} \Sigma_{ij} V_{j\ell} = \sum_{i=1}^{N_A} \Sigma_{ii} U_{ki} V_{i\ell}.$$

Hence,

$$|\psi\rangle = \sum_{k,\ell} \sum_{i} \Sigma_{ii} U_{ki} |\phi_k\rangle \otimes V_{i\ell} |\chi_\ell\rangle = \sum_{i} \Sigma_{ii} \underbrace{\left(\sum_{k} U_{ki} |\phi_k\rangle \right)}_{|e_i\rangle} \otimes \underbrace{\left(\sum_{\ell} V_{j\ell} |\chi_\ell\rangle \right)}_{|j_B\rangle}.$$

□

Proposition: Schmidt Rank Entanglement Criterion

Proposition 1.30 $|\psi\rangle$ is entangled iff its Schmidt rank is > 1. Otherwise, it is separable (i.e. a product state).

Definition: State.Maximally Entangled

Definition 1.31 Let $|\psi\rangle$ be a pure state in a bipartite system $\mathbb{H}_{AB} = \mathbb{H}_A \otimes \mathbb{H}_B$, where $\mathbb{H}_A$ has dimension N_A and $\mathbb{H}_B$ has dimension $N_B \geq N_A$. $|\psi\rangle$ is **maximally entangled** if all its Schmidt coefficients are equal (to $1/\sqrt{N_A}$).

Notation 1.32 Write $\mathbb{S}(\mathbb{H}) = \left\{\rho \in \mathbb{B}(\mathbb{H}) : \rho = \rho^\dagger, \rho \geq 0, \operatorname{tr} \rho = 1\right\}$ for the set of density matrices on $\mathbb{H}$.

Definition: Partial Trace

Definition 1.33 The **partial trace** over B, tr_B, on the bipartite system $\mathbb{H}_{AB} = \mathbb{H}_A \otimes \mathbb{H}_B$ is the operator defined linearly by

$$\mathrm{tr}_B : \mathbb{S}(\mathbb{H}_{AB}) \to \mathbb{S}(\mathbb{H}_A),$$
$$|a_1\rangle\langle a_2| \otimes |b_1\rangle\langle b_2| \mapsto \mathrm{tr}(|b_1\rangle\langle b_2|) \cdot |a_1\rangle\langle a_2|.$$

Note that if $\rho_{AB} = \rho_A \otimes \rho_B$, then $\mathrm{tr}_B \rho_{AB} = \mathrm{tr}(\rho_B) \cdot \rho_A = \rho_A$.

Definition: Reduced Density Matrix

Definition 1.34 Let ρ_{AB} be a density matrix in $\mathbb{S}(\mathbb{H}_{AB})$. $\rho_A = \mathrm{tr}_B(\rho_{AB})$ is called the **reduced density matrix** or **marginal** of ρ_{AB} in A

Proposition: Equivalent Definition Of Partial Trace

Proposition 1.35 Let $M_A \in B(\mathbb{H}_A)$. We have

$$\mathrm{tr}(M_A \rho_A) = \mathrm{tr}((M_A \otimes \mathbb{I}_B)\rho_{AB}).$$

for all $\rho_{AB} \in \mathbb{S}(\mathbb{H}_{AB})$, $\rho_A = \mathrm{tr}_B(\rho_{AB})$. In fact, this can be taken to be an equivalent definition of partial trace.

Remark 1.36 Let $\rho_{AB} = |\psi\rangle\langle\psi| \in \mathbb{S}(\mathbb{H}_{AB})$ be a pure state and let r_ψ be its Schmidt rank. Then

$$\rho_A = \mathrm{tr}_B(|\psi\rangle\langle\psi|) = \sum_{k=1}^{r_\psi} p_k |u_k\rangle\langle u_k|.$$

So ρ_A is pure iff $r_\psi = 1$, i.e. iff $|\psi\rangle$ is separable.

Proposition: Partial Trace Properties

Proposition 1.37 Let $\rho_{AB} \in B(\mathbb{H}_{AB})$ and $\rho_A = \mathrm{tr}_B(\rho_{AB})$. Then:

1. $\mathrm{tr}\,\rho_A = \mathrm{tr}\,\rho_{AB}$.
2. If $\rho_{AB} \geq 0$, then $\rho_A \geq 0$.
3. If ρ_{AB} is a density matrix then ρ_A is a density matrix.
4. We have

$$\langle \phi_i | \rho_A | \phi_i \rangle = \sum_k \langle \phi_i \otimes \psi_k | \rho_{AB} | \phi_i \otimes \psi_k \rangle,$$

   for an orthonormal bases $\{|\phi_i\rangle\}$ and $\{|\psi_k\rangle\}$.
5. If $\rho_{AB} = \sigma_A \otimes \sigma_B$ and $\mathrm{tr}(\sigma_B) = 1$, then $\sigma_A = \rho_A$.

Proof (Hints).

1. Straightforward.
2. Use Proposition 1.35.
3. Straightforward.

□

Proof.

1. This follows from linearity of trace and the fact that $\operatorname{tr}(\rho \otimes \sigma) = \operatorname{tr}(\rho) \cdot \operatorname{tr}(\sigma)$.
2. By Proposition 1.35,

$$\begin{aligned}\langle\psi|\rho_A|\psi\rangle &= \operatorname{tr}(\rho_A|\psi\rangle\langle\psi|) = \operatorname{tr}(\rho_{AB}(|\psi\rangle\langle\psi| \otimes \mathbb{I})) \\ &= \sum_i \operatorname{tr}(\rho_{AB}|\psi\rangle\langle\psi| \otimes |i\rangle\langle i|) = \sum_i (\langle\psi| \otimes \langle i|)\rho_{AB}(|\psi\rangle \otimes |i\rangle) \geq 0.\end{aligned}$$

3. From 1 and 2, by definition.

□

Definition: Purification

Definition 1.38 Let $\rho_A \in \mathbb{S}(H_A)$ be a (pure or mixed) state. We may introduce an auxiliary space $\mathbb{H}_R$ of dimension $\mathrm{rank}(\rho_A)$ and construct a pure state $|\psi_{AR}\rangle \in \mathbb{H}_A \otimes \mathbb{H}_R$ such that $\rho_A = \mathrm{tr}_R(|\psi_{AR}\rangle\langle\psi_{AR}|)$. This is called **purification**.

Remark 1.39 Let $\left\{M_n^A\right\}_n$ be a POVM in $\mathbb{H}_A$. Then $\left\{M_n^A \otimes \mathbb{I}_B\right\}_n$ is a POVM in $\mathbb{H}_{AB}$.

Theorem: Naimark

Theorem 1.40 (Naimark) For every POVM $\{E_n\}_{n=1}^m \subseteq \mathbb{B}(\mathbb{H})$, there is a state $|\psi\rangle \in \mathbb{C}^m$ and a projective measurement $\{P_n\}_{n=1}^m \subseteq B(\mathbb{H} \otimes \mathbb{C}^m)$ such that

$$\mathrm{tr}(\rho E_n) = \mathrm{tr}((\rho \otimes |\psi\rangle\langle\psi|)P_n) \quad \forall n \in [m], \forall \rho \in \mathbb{S}(\mathbb{H}).$$

2. Quantum channels and open systems

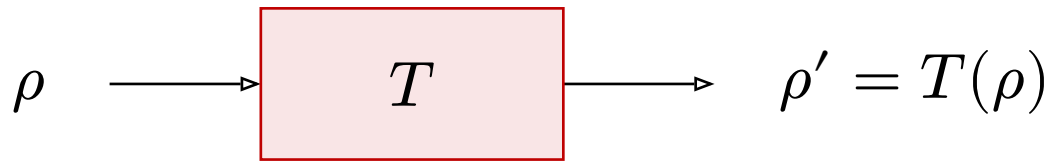
2.1. Quantum channels

Definition: Quantum Channel

Definition 2.1 A **quantum channel** is a linear map $T : \mathbb{S}(\mathbb{H}_{\text{in}}) \to \mathbb{S}(\mathbb{H}_{\text{out}})$ which satisfies:

- **Preserves trace**: $\text{tr}(T(\rho)) = \text{tr}(\rho)$ for all $\rho \in \mathbb{S}(\mathbb{H}_{\text{in}})$.
- **Positive**: if $\rho \geq 0$, then $T(\rho) \geq 0$.
- **Completely positive**: for all $n \in \mathbb{N}$, $(T \otimes \mathbb{I}_n)$ is positive (note that this implies the second condition, but the converse is false).

So quantum channels are completely positive trace-preserving (CPTP) maps. We may depict a quantum channel T as follows:


ρ
T
$\rho' = T(\rho)$

Example 2.2 Examples of quantum channels:

- Unitary evolution: $\rho \mapsto U\rho U^*$.
- Adding an ancilla: $\rho \mapsto \rho \otimes \rho_E$ (the E denotes "environment").
- Partial trace: $\rho \mapsto \mathrm{tr}_B(\rho)$ or $\rho \mapsto \mathrm{tr}_A(\rho)$.

We will see that in fact, any quantum channel is a combination of these three.

Definition: Maximally Entangled State

Definition 2.3 We define the **maximally entangled state** in $\left(\mathbb{C}^d\right)^{\otimes 2}$ as

$$|\phi\rangle = \frac{1}{\sqrt{d}} \sum_{k=1}^{d} |kk\rangle.$$

Definition: Partial Transpose

Definition 2.4 Recall the transposition map is defined as

$$\Theta : A \to A^T, \quad \langle i|A^T|j\rangle = \langle j|A|i\rangle.$$

We define the **partial transpose** by its action on the maximally entangled state $|\phi\rangle = \frac{1}{d}\sum_{i=1}^{d}|ii\rangle$:

$$(|\phi\rangle\langle\phi|)^{T_A} = (|\phi\rangle\langle\phi|)^{T_1} = (\Theta \otimes \mathrm{id})(|\phi\rangle\langle\phi|) = \frac{1}{d}F,$$

where $F = \sum_{i,j=1}^{n}|ij\rangle\langle ji|$ is the flip operator. Note the partial transpose is positive but not CP. Alternatively, we can define it by its action on an orthonormal basis:

$$\langle ij|X^{T_A}|k\ell\rangle = \langle kj|X|i\ell\rangle.$$

Remark 2.5 Note that the partial transpose is useful for detecting entanglement but is not physically implementable (as not CP).

Definition: Hilbert Schmidt Inner Product

Definition 2.6 The **Hilbert-Schmidt inner product** of $A, B \in B\big(\mathbb{C}^d\big)$ is

$$\langle A|B\rangle_{\mathrm{HS}} := \mathrm{tr}(A^* B).$$

Definition: Superoperator Adjoint

Definition 2.7 Let $T : B\left(\mathbb{C}^d\right) \to B\left(\mathbb{C}^{d'}\right)$ be a linear map. The **adjoint** of T is the unique linear map $T^* : B\left(\mathbb{C}^{d'}\right) \to B\left(\mathbb{C}^d\right)$ which satisfies

$$\langle T(A), B\rangle_{\mathrm{HS}} = \langle A, T^*(B)\rangle_{\mathrm{HS}} \quad \forall A \in B\left(\mathbb{C}^d\right), B \in B\left(\mathbb{C}^{d'}\right).$$

Definition: Choi Jamiolkowski Matrix

Definition 2.8 Let $T : B\left(\mathbb{C}^d\right) \to B\left(\mathbb{C}^{d'}\right)$ be a linear map. The **Choi-Jamiolkowski matrix** $C \in B\left(\mathbb{C}^{d'} \otimes \mathbb{C}^d\right)$ of T is defined as

$$C := (T \otimes \mathrm{id}_d)|\phi\rangle\langle\phi|.$$

Note that in fact, $C \in S\left(\mathbb{C}^{d'} \otimes \mathbb{C}^d\right)$ is a density matrix if T is a quantum channel.

Remark 2.9 Note that the Choi-Jamiolkowski matrix completely determines T: since $|\phi\rangle\langle\phi| = \frac{1}{d}\sum_{n,m=1}^{d}|nn\rangle\langle mm|$, we have

$$\begin{aligned}\langle ij|C|k\ell\rangle &= \frac{1}{d}\sum_{m,n=1}^{d}\langle ij|(T(|n\rangle\langle m|)\otimes|n\rangle\langle m|)|k\ell\rangle \\ &= \frac{1}{d}\sum_{m,n=1}^{d}\langle j|n\rangle\cdot\langle m|\ell\rangle\cdot\langle i|T(|n\rangle\langle m|)|k\rangle = \frac{1}{d}\langle i|T(|j\rangle\langle\ell|)|k\rangle,\end{aligned}$$

and so we can determine any entry of any $T(\rho)$ by linearity. This state-channel duality is called the **Choi-Jamiolkowski isomorphism**, and can be expressed as

$$\operatorname{tr}(AT(B)) = d \operatorname{tr}\big(C\big(A \otimes B^T\big)\big) \quad \forall A \in B\Big(\mathbb{C}^{d'}\Big), B \in B\big(\mathbb{C}^d\big).$$

Indeed, let $\mathbb{F}|ij\rangle = |ji\rangle$ be the flip operator: note that $\mathbb{F}^{T_2} = d|\phi\rangle\langle\phi|$, then if $d = d'$,

$$\begin{aligned} d \operatorname{tr}\big(C\big(A \otimes B^T\big)\big) &= d \operatorname{tr}\big((T \otimes \mathrm{id}_d)(|\phi\rangle\langle\phi|)\big(A \otimes B^T\big)\big) \\ &= \operatorname{tr}\big((T \otimes \mathrm{id}_d)\mathbb{F}^{T_2}\big(A \otimes B^T\big)\big) \\ &= \operatorname{tr}\big(\mathbb{F}^{T_2}\big(T^*(A) \otimes B^T\big)\big) = \operatorname{tr}(T^*(A) \otimes B) = \operatorname{tr}(AT(B)). \end{aligned}$$

Theorem: Characterisation Of Quantum Channels

Theorem 2.10 (Characterisation of Quantum Channels) Let $T : B\big(\mathbb{C}^d\big) \to B\Big(\mathbb{C}^{d'}\Big)$ be a linear map. TFAE:

1. T is a quantum channel.
2. Let $C = (T \otimes \mathbb{I}_d)(|\phi\rangle\langle\phi|)$ be the Choi-Jamiolkowski matrix of T, then $C \geq 0$ and $\mathrm{tr}_1(C) = \frac{1}{d}\mathbb{I}_d$.
3. **Kraus decomposition**: There exists $\{A_k\}_{k=1}^{dd'} \subseteq \mathbb{C}^{d' \times d}$ with $\sum_{k=1}^{dd'} A_k^* A_k = \mathbb{I}_d$ such that

$$T(\rho) = \sum_{k=1}^{dd'} A_k \rho A_k^* \quad \forall \rho \in S\big(\mathbb{C}^d\big).$$

We call the number of non-trivial A_k in the Kraus decomposition the **Kraus rank** of T.

4. **Stinespring dilation**: there exists a unitary U on $\mathbb{C}^d \otimes \mathbb{C}^{dd'}$ and a state $|\psi\rangle \in \mathbb{C}^{dd'}$ such that $T(\rho) = \mathrm{tr}_{\mathbf{2}}(U(\rho \otimes |\psi\rangle\langle\psi|)U^*)$ for all $\rho \in S\big(\mathbb{C}^d\big)$.

Proof (Hints).

- $1 \Rightarrow 2$: straightforward.
- $4 \Rightarrow 1$: use that compositions of quantum channels are quantum channels.

□

Proof.

- $1 \Rightarrow 2$: $C \geq 0$ follows from the completely positive property of T and linearity. Also,

$$\begin{aligned}
\mathrm{tr}_1(C) &= \frac{1}{d} \sum_{n,m=1}^{d} \mathrm{tr}(T|n\rangle\langle m|) \cdot |n\rangle\langle m| \\
&= \frac{1}{d} \sum_{n,m=1}^{d} \mathrm{tr}(|n\rangle\langle m|) \cdot |n\rangle\langle m| \quad \text{since } T \text{ preserves trace} \\
&= \frac{1}{d} \sum_{n,m} \delta_{mn} |n\rangle\langle m| = \frac{1}{d} \sum_{n=1}^{d} |n\rangle\langle n| = \frac{1}{d} \mathbb{I}_d.
\end{aligned}$$

- $2 \Rightarrow 3$: we use that (verify this) $(A \otimes \mathbb{I})|\phi\rangle = (\mathbb{I} \otimes A^T)|\phi\rangle$ for all $A \in B(\mathbb{C}^d)$, where $|\phi\rangle$ is the maximally entangled state, and that $\forall |\psi\rangle \in \mathbb{C}^{d^2}$, there exists A such that $|\psi\rangle = (A \otimes \mathbb{I})|\phi\rangle$. Since $C \geq 0$, we can write $C = \sum_{k=1}^{dd'} |\psi_k\rangle\langle\psi_k|$ ($|\psi_k\rangle$ are not necessarily normalised). So

$$\begin{aligned} C &= \sum_{k=1}^{dd'} (A_k \otimes \mathbb{I})|\phi\rangle\langle\phi|(A_k^* \otimes \mathbb{I}) \\ &= (T \otimes \mathbb{I})|\phi\rangle\langle\phi|. \end{aligned}$$

Also,

$$
\begin{aligned}
\frac{1}{d}\mathbb{I} &= \mathrm{tr}_1(C) = \sum_{n=1}^{d} \langle n_1 | C_{12} | n_1 \rangle \\
&= \frac{1}{d} \sum_{n=1}^{d} \sum_{m=1}^{dd'} (\mathbb{I} \otimes A_m^T)(|\phi\rangle\langle\phi|) \left(\mathbb{I} \otimes \overline{A}_k\right) |n\rangle \\
&= \sum_{n=1}^{d} \langle n | \sum_{k=1}^{dd'} (\mathbb{I} \otimes A_m^T) \frac{1}{d} \left(\sum_{k,\ell=1}^{d} |kk\rangle\langle\ell\ell| \right) \left(\mathbb{I} \otimes \overline{A}_k\right) |n\rangle \\
&= \frac{1}{d} \sum_{n=1}^{d} \sum_{m=1}^{dd'} \sum_{k,\ell=1}^{d} \langle n|k\rangle\langle\ell|n\rangle A_m^T |k\rangle\langle\ell| \overline{A}_k
\end{aligned}
$$

$$= \frac{1}{d} \sum_{n=1}^{d} \sum_{m=1}^{dd'} A_m^T |n\rangle\langle n| \overline{A}_m$$

$$= \frac{1}{d} \sum_{m=1}^{dd'} A_m^T \overline{A}_m$$

So we set $\tilde{A}_m := \overline{A}_m$.

- $3 \Rightarrow 4$: let $V = \sum_{k=1}^{dd'} A_k \otimes |k\rangle$, where $\{|k\rangle\}_{k=1}^{dd'}$ is an orthonormal basis of $\mathbb{C}^{dd'}$. V is an isometry, i.e. $V^*V = \sum_{k=1}^{dd'} A_k^* A_k = \mathbb{I}_d$. Then for all $\rho \in S\left(\mathbb{C}^{dd'}\right)$, since $(A_k \otimes |k\rangle)\rho = (A_k \rho) \otimes |k\rangle$,

$$\begin{aligned}\mathrm{tr}_2(V\rho V^*) &= \mathrm{tr}_2\left(\sum_{k,\ell=1}^{dd'}(A_k\rho A_\ell^*)\otimes|k\rangle\langle\ell|\right)\\ &= \sum_{k,\ell=1}^{dd'}(A_k\rho A_\ell^*)\,\mathrm{tr}(|k\rangle\langle\ell|)\\ &= \sum_{k=1}^{dd'}A_k\rho A_k^* = T(\rho).\end{aligned}$$

Now choose $V = U(\mathbb{I}\otimes|\psi\rangle)$ for some pure state $|\psi\rangle$ and unitary U.

- $4 \Rightarrow 1$: the maps

$$\rho \mapsto \rho \otimes |\psi\rangle\langle\psi| \mapsto U(\rho \otimes |\psi\rangle\langle\psi|)U^* \mapsto \mathrm{tr}_2(U(\rho \otimes |\psi\rangle\langle\psi|)U^*)$$

are all quantum channels, and so their composition is also a quantum channel.

□

Remark 2.11

- The number k in the Kraus decomposition is called the **Kraus rank** of T, which is the same as the Choi rank (rank of the Choi-Jamiolkowski matrix). Note: this is not the same as the rank of T as a map.
- We can always express T with $r = \text{rank}(C)$ Kraus operators which are orthogonal (w.r.t Hilbert-Schmidt inner product), since T is a completely positive linear map.
- Two sets of Kraus operator $\{K_j\}$ and $\{J_\ell\}$ represent the same map T iff there exists a unitary U such that $K_j = \sum_\ell U_{j\ell} J_\ell$.

2.2. Examples of quantum channels

Definition: Depolarising Channel

Definition 2.12 In two dimensions, there are three kinds of errors:

1. Bit flip errors, modelled by the Pauli X: $|0\rangle \mapsto |1\rangle$, $|1\rangle \mapsto |0\rangle$.
2. Phase flip error: modelled by Pauli Z: $|0\rangle \mapsto |0\rangle$, $|1\rangle \mapsto -|1\rangle$.
3. Combination of bit and phase flip errors: modelled by Pauli Y.

A unitary map describing the depolarising channel is

$$_{AE} : |\psi\rangle_A \mapsto \sqrt{1-p}|\psi\rangle_A \otimes |0\rangle_E + \sqrt{\frac{p}{3}}(X|\psi\rangle_A \otimes |1\rangle_E + Y|\psi\rangle_A \otimes |2\rangle_E + Z|\psi\rangle_A \otimes$$

(the environment H_E has dimension 4). We can express this in the Kraus decomposition: let $M_a := \langle a|_E U_{A\to AE}$, $a \in \{0,1,2,3\}$, and $M_0 =$

$\sqrt{1-p}\mathbb{I}$, $M_1 = \sqrt{p/3}X$, $M_2 = \sqrt{p/3}Y$, $M_3 = \sqrt{p/3}Z$. It is straightforward to see that

$$\sum_{a=0}^{3} M_a^\dagger M_a = \left(1 - p + \frac{p}{3} + \frac{p}{3} + \frac{p}{3}\right)\mathbb{I} = \mathbb{I}.$$

The channel is (tracing out the ancilla) $T(\rho) = (1-p)\rho + \frac{p}{3}(X\rho X + Y\rho Y + Z\rho Z)$. For arbitrary dimensions D, the depolarising channel is $\rho \mapsto (1-p)\rho + p\sigma$, where $\sigma \in S(\mathbb{C}^D)$, usually $\sigma = \mathbb{I}/d$.

Definition: Phase Damping Channel

Definition 2.13 The **phase damping channel** is the map

$$\rho = \begin{bmatrix} \rho_{00} & \rho_{01} \\ \rho_{10} & \rho_{11} \end{bmatrix} \mapsto \begin{bmatrix} \rho_{00} & (1-p)\rho_{01} \\ (1-p)\rho_{10} & \rho_{11} \end{bmatrix}.$$

Let the environment have orthonormal basis $\{|0\rangle, |1\rangle, |2\rangle\}$, then the state representation is

$$|0\rangle_A \mapsto \sqrt{1-p}|0\rangle_A \otimes |0\rangle_E + \sqrt{p}|0\rangle_A \otimes |1\rangle_E$$
$$|1\rangle_A \mapsto \sqrt{1-p}|1\rangle_A \otimes |0\rangle_E + \sqrt{p}|1\rangle_A \otimes |2\rangle_E$$

The Kraus operators are $M_0 = \sqrt{1-p}\cdot\mathbb{I}$, $M_1 = \sqrt{p}|0\rangle\langle 0|$, $M_2 = \sqrt{p}|1\rangle\langle 1|$. We have $M_0^2 + M_1^2 + M_2^2 = \mathbb{I}$. The map is (tracing out the ancilla) $T(\rho) = (1-p/2)\rho + \frac{1}{2}pZ\rho Z$.

Definition: Separable Density Matrix

Definition 2.14 A density matrix $\rho \in \mathbb{S}(\mathbb{H}_A \otimes \mathbb{H}_B)$ is **separable** if it can be expressed as a convex combination

$$\rho = \sum_i p_i \rho_i^A \otimes \sigma_i^B,$$

where $p_i \geq 0$, $\sum_i p_i = 1$, and $\rho_i^A \in \mathbb{S}(\mathbb{H}_A)$ and $\sigma_i^B \in \mathbb{S}(\mathbb{H}_B)$.

Definition: Entanglement Breaking

Definition 2.15 A quantum channel T is **entanglement breaking** if its Choi-Jamiolkowski matrix is separable. This is equivalent to the existence of a POVM $\{M_k\}$ and a set of density matrices $\{\rho_k\}$ such that $T(\rho) = \sum_k \mathrm{tr}(M_k \rho)\rho_k$.

2.3. Properties of channels

Remark 2.16 Let $|\psi\rangle \in \mathbb{H}_A \otimes \mathbb{H}_B$, $d = \min\{\dim H_A, \dim H_B\}$, not necessarily normalised. The Schmidt decomposition is

$$|\psi\rangle = \sum_{j=1}^{d} \lambda_j |e_j\rangle \otimes |f_j\rangle,$$

where $\lambda_j \geq 0$, $\sum_{j=1}^{d} \lambda_j^2 = \langle\psi|\psi\rangle$, and $\{|e_j\rangle : j \in [d]\}$ and $\{|f_j\rangle : j \in [d]\}$ orthonormal bases.

The reduced density operators of $|\psi\rangle$ are diagonal in the bases $\{|e_j\rangle\}$, $\{|f_j\rangle\}$, with eigenvalues λ_j^2. Conversely, if $\rho_A \in \mathbb{S}(\mathbb{H}_A)$ has spectral decomposition $\rho_A = \sum_j \lambda_j^2 |e_j\rangle\langle e_j|$, then $|\psi\rangle$ provides a purification for

$\rho_A = \mathrm{tr}_B(|\psi\rangle\langle\psi|)$; the minimal dilation space we can choose, $\mathbb{H}_{\min}$ has dimension $\mathbf{rank}(\rho_A)$. If $|\psi\rangle \in \mathbb{H}_A \otimes \mathbb{H}_{\min}$, then all other purifications of ρ_A are of the form $|\psi'\rangle = (\mathbb{I}_A \otimes V)|\psi\rangle$, with $V \in B(\mathbb{H}_{\min}, \mathbb{H}_B)$ an isometry. Hence, all purifications are related by $\mathbb{I}_A \otimes U$ with U an isometry.

Proposition: Equivalence Of Ensembles

Proposition 2.17 (Equivalence of Ensembles) Let $\{|\psi_j\rangle : j \in [M]\}$ and $\{|\phi_\ell\rangle : \ell \in [N]\}$ be (not necessarily normalised) ensembles. Then

$$\sum_{j=1}^{M} |\psi_j\rangle\langle\psi_j| = \sum_{\ell=1}^{N} |\phi_\ell\rangle\langle\phi_\ell|$$

iff there is an isometry $U \in \mathbb{C}^{M\times N}$ such that $|\psi_j\rangle = \sum_{\ell=1}^{N} U_{j\ell}|\phi_\ell\rangle$.

Proof (Hints).

- $\Longleftarrow$: straightforward.
- $\Longrightarrow$: explain why we can assume that $\rho = \sum_j |\psi_j\rangle\langle\psi_j|$ and $\sigma = \sum_\ell |\phi_\ell\rangle\langle\phi_\ell|$ are density matrices. Consider purifications of ρ and σ which use the computational basis in the dilation space.

□

Proof.

- $\Longleftarrow$: this is straightforward to show.
- $\Longrightarrow$: WLOG (by rescaling ρ), we can assume $\rho := \sum_j |\psi_j\rangle\langle\psi_j|$ is a density matrix. We have $\rho = \mathrm{tr}_B(|\psi\rangle\langle\psi|)$ (through purification), where $|\psi\rangle = \sum_j |\psi_j\rangle \otimes |j\rangle$. Similarly, let $|\phi\rangle = \sum_\ell |\phi_\ell\rangle \otimes |\ell\rangle$ (so we use the same orthonormal basis $\{|\ell\rangle\} = \{|j\rangle\}$). So $|\psi\rangle$ and $|\phi\rangle$ differ by a unitary (or an isometry if the dimensions are not equal), hence $|\psi\rangle = (1 \otimes U)|\phi\rangle$. Taking the scalar product with $\langle j|$, we obtain $|\psi_j\rangle = \sum_\ell U_{j\ell}|\phi_\ell\rangle$.

□

Notation 2.18 Let T_1, T_2 be linear maps. Write $T_2 \geq T_1$ to mean $T_2 - T_1$ is completely positive. By the Choi-Jamiolkowski isomorphism, this is equivalent to $C_2 \geq C_1$ where C_i is the Choi matrix of T_i (i.e. $C_2 - C_1$ is positive semi-definite): if $T_2 - T_1$ is completely positive, then

$$C_2 - C_1 = ((T_2 - T_1) \otimes \mathbb{I}_d)(|\phi\rangle\langle\phi|) \geq 0,$$

since $|\phi\rangle\langle\phi|$ is positive.

Theorem 2.19 Let $T_1, T_2 : \mathbb{C}^{d' \times d'} \to \mathbb{C}^{d \times d}$ be completely positive maps, with $T_2 \geq T_1$. Let $V_i : \mathbb{C}^d \to \mathbb{C}^{d'} \otimes \mathbb{C}^{r_i}$ be Stinespring representations for T_i (i.e. $T_i(A) = V_i^* \Big(A \otimes \mathbb{I}_{r_i} \Big) V_i$), then there is a contraction (i.e. $W^*W \leq \mathbb{I}$) $W : \mathbb{C}^{r_2} \to \mathbb{C}^{r_1}$ such that $V_1 = (\mathbb{I}_{d'} \otimes W) V_2$.

Moreover, if V_2 belongs to a minimal dilation, then W is unique.

Proof (Hints).

□

Proof. We use the equivalence $T_2 \geq T_1 \Leftrightarrow C_2 \geq C_1$. Define the map

$$R_i = \Big(\mathbb{I}_{r_i} \otimes \langle\phi|\Big)(V_i \otimes \mathbb{I}_{d'}) \in B\Big(\mathbb{C}^d \otimes \mathbb{C}^{d'}, \mathbb{C}^{r_i}\Big)$$

Let $|\psi\rangle \in \mathbb{C}^d \otimes \mathbb{C}^{d'}$. We want to show $\|R_2|\psi\rangle\|^2 \geq \|R_1|\psi\rangle\|^2$. Indeed,

$$\begin{aligned}
\|R_2|\psi\rangle\|^2 &= \langle\psi|R_2^* R_2|\psi\rangle \\
&= \langle\psi|(V_2^* \otimes \mathbb{I}_{d'})\Big(\mathbb{I}_{r_2} \otimes |\phi\rangle\Big)\Big(\mathbb{I}_{r_2} \otimes \langle\phi|\Big)(V_2 \otimes \mathbb{I}_{d'})|\psi\rangle \\
&= \langle\psi|(T_2 \otimes \mathrm{id})(|\phi\rangle\langle\phi|)\rangle \\
&= \langle\psi|C_2|\psi\rangle \geq \langle\psi|C_1|\psi\rangle.
\end{aligned}$$

And $\langle\psi|C_1|\psi\rangle = \|R_1|\psi\rangle\|^2$ by the same argument. So there exists a contraction $W : \mathbb{C}^{r_2} \to \mathbb{C}^{r_1}$, such that $R_1 = WR_2$. So $V_1 = (\mathbb{I}_{d'} \otimes W)V_2$. If $r_2 = \mathrm{rank}(C_2)$, then R_2 is surjective, and so W is uniquely determined. $\square$

Theorem: Radon Nikodym

Theorem 2.20 (Radon-Nikodym) Let $\{T_i\}$ be a set of CP maps such that $\sum_i T_i = T \in B\left(\mathbb{C}^{d' \times d'}, \mathbb{C}^{d \times d}\right)$ with Stinespring representation $T(A) = V^*(A \otimes \mathbb{I}_r)V$. Then there exists a set of non-negative operators $P_i \in \mathbb{C}^{r \times r}$ such that $\sum_i P_i = \mathbb{I}_r$ and $T_i(A) = V^*(A \otimes P_i)V$.

Remark 2.21 Since $T = \sum_i T_i$, this gives $T(A) = \sum_i V^*(A \otimes P_i)V$, where $\{P_i\}$ is a POVM. This gives an identification between quantum channels of this form and POVMs.

Definition: Instrument

Definition 2.22 An **instrument** is a set of CP maps $\{T_i\}$ whose sum is trace-preserving.

Remark 2.23 Instruments encompass the notions of quantum channels and POVMs:

- We can assing a quantum channel $T : \rho \mapsto \sum_i T_i(\rho)$. (Measurement outcome ignored.)
- By contrast, POVMs ignore the quantum system: $p_i = \mathrm{tr}(T_i(\rho)) = \mathrm{tr}(T_i(\rho)\mathbb{I}) = \mathrm{tr}(\rho T_i^*(\mathbb{I})) =: \mathrm{tr}(\rho M_i)$: $\{M_i\}$ is a POVM.

Remark 2.24 Instruments can viewed as a special case of quantum channels by assigning to them the quantum channel

$$\rho \mapsto \sum_i T_i(\rho) \otimes |i\rangle\langle i|,$$

where $\{|i\rangle\}$ is an orthonormal basis.

Proposition: Quantum Steering

Proposition 2.25 (Quantum Steering) Let $\rho \in B(\mathbb{H}_A)$ be a density operator with purification $|\psi\rangle \in \mathbb{H}_A \otimes \mathbb{H}_B$. Let $\rho = \sum_i \lambda_i \rho_i$ be a convex combination. Then there is an instrument $\{T_i\}$ with each $T_i : B(\mathbb{H}_B) \to B(\mathbb{H}_B)$, such that $\lambda_i \rho_i = \mathrm{tr}_B((\mathbb{I} \otimes T_i)(|\psi\rangle\langle\psi|))$.

2.4. Description of open quantum many-body systems

Assume evolution is

$$\rho_{SE}(t) = \rho_S(t) \otimes \rho_E \overset{\mathrm{d}t}{\mapsto} \rho_{SE}(t+\mathrm{d}t) = \rho_S(t+\mathrm{d}t) \otimes \rho_E(t+\mathrm{d}t) = \rho_S(t+\mathrm{d}t) \otimes \rho_E$$

Definition: Quantum Markov Semigroup

Definition 2.26 A **quantum Markov semigroup** is a 1-parameter continuous semigroup $\{T_t : t \geq 0\}$ of quantum channels (so each $T_t : \mathbb{S}(\mathbb{H}) \to \mathbb{S}(\mathbb{H})$).

Note that $T_0 = \mathbb{I}$ and $T_s \circ T_t = T_{t+s}$. We have

$$\frac{\mathrm{d}}{\mathrm{d}t} T_t = \mathcal{L} \circ T_t = T_t \circ \mathcal{L},$$

where $\mathcal{L}$ is the infinitesimal generator of the semigroup, called the **Liouvillian** or **Lindbladian**. This equation is called the **master equation** or **Liouville equation**. This gives

$$T_t = e^{t\mathcal{L}}.$$

Definition: Markovian Quantum Master Equation

Definition 2.27 The **Markovian quantum master equation** is

$$\mathcal{L}(\rho) = -i[H, \rho] + \sum_j \gamma_j \left(A_j \rho A_j^* - \frac{1}{2}\{A_j^* A_j, \rho\} \right),$$

where the A_j are called the **jump operators** and the γ_j are usually taken to be 1. The **quantum master equation** corresponding to this Lindbladian is

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho(t) = \mathcal{L}(\rho(t)).$$

This gives $\rho(t) = e^{\mathcal{L}t}\rho(0) = T_t(\rho(0))$.

2.5. Separability criteria

Notation 2.28 Let $A(\mathbb{H})$ denote the set of bounded linear Hermitian operators on $\mathbb{H}$.

Definition: Operator Correlation

Definition 2.29 The **covariance** (or **operator correlation**) of ρ between subsystems A and B is

$$\mathrm{Cor}_\rho(A:B) = \sup_{\|M_A\|,\|M_B\|\leq 1} |\mathrm{tr}(\rho M_A T_B) - \mathrm{tr}(\rho M_A)\,\mathrm{tr}(\rho M_B)|,$$

where $M_A \in A(H_A)$, $M_B \in A(H_B)$, and $\|\cdot\|$ is the standard operator norm.

Example 2.30 If ρ is separable, then $\mathrm{Cor}_\rho(A:B)$ measures classical correlation. If $\rho = \rho_A \otimes \rho_B$, then $\mathrm{Cor}_\rho(A:B) = 0$.

Definition: Entanglement Entropy

Definition 2.31 Let $|\psi\rangle = \sum_{i=1}^{d} \sqrt{p_i}|e_i\rangle \otimes |f_i\rangle$ be the Schmidt decomposition of $|\psi\rangle \in \mathbb{H}_A \otimes \mathbb{H}_B$. Let $\rho = |\psi\rangle\langle\psi|$. The **entanglement entropy** of ρ is the Shannon entropy of the probability distribution $(p_1, ..., p_d)$:

$$S_{\text{ENT}}(\rho) := -\sum_{i=1}^{d} p_i \log(p_i).$$

Proposition: Properties Of Entanglement Entropy

Proposition 2.32

- $S_{\mathrm{ENT}(\rho)} = 0$ iff the Schmidt rank of $|\psi\rangle$ is 1.
- The maximum value of $S_{\mathrm{ENT}}(\rho)$ is $\log(d)$, and is achieved iff $|\psi\rangle$ is maximally entangled, i.e. $\lambda_i = 1/d$ for all $i \in [d]$.

Proposition: Ppt Criterion

Proposition 2.33 (PPT Criterion) Let $\rho \in \mathbb{S}(\mathbb{H}_A \otimes \mathbb{H}_B)$. If ρ^{T_A} has a negative eigenvalue, then ρ is entangled.

Proof (Hints). Prove the contrapositive. □

Proof. Assume ρ is separable, so $\rho = \sum_j p_j \rho_j^A \otimes \rho_j^B$. Then

$$\rho^{T_A} = (\Theta \otimes \mathrm{id})(\rho) = \sum_j p_j \left(\rho_j^A\right)^T \otimes \rho_j^B,$$

and so $\rho^{T_A} \geq 0$, as it is a sum of positive matrices. $\square$

Definition: Entanglement Witness

Definition 2.34 Write $S_{\text{SEP}} = \{\text{separable density matrices}\}$, which is convex and compact. By the Hahn-Banach theorem, for all $\rho \notin S_{\text{SEP}}$, there exists a hyperplane determined by a Hermitian operator ω such that $\text{tr}(\rho\omega) < 0$ and $\text{tr}(\sigma\omega) \geq 0$ for all $\sigma \in S_{\text{SEP}}$. ω is called an **entanglement witness** for ρ.

By the Choi-Jamiolkowski isomorphism, ω corresponds to a map Λ via the following:

$$\omega = (\Lambda \otimes \text{id}_B)(|\phi\rangle\langle\phi|).$$

Remark 2.35 The entanglement witness corresponding to the transposition map is the flip operator F.

Proposition: Positive Map Criterion For Separability

Proposition 2.36 Let $\mathbb{H}_{AB} = \mathbb{H}_A \otimes \mathbb{H}_B$ and let $\rho \in \mathbb{S}(\mathbb{H}_{AB})$. Then ρ is separable iff $(\Lambda \otimes \mathrm{id}_B)(\rho) \geq 0$ for every positive map $\Lambda : B(\mathbb{H}_A) \to B(\mathbb{H}_A)$.

Proof (Hints).

- $\Longrightarrow$: straightforward.
- $\Longleftarrow$: TODO.

□

Proof. $\implies$: let ρ be separable, so we can write $\rho = \sum_j p_j \rho_j \otimes \sigma_j$. Then for every positive $\Lambda : B(\mathbb{H}_A) \to B(\mathbb{H}_A)$,

$$(\Lambda \otimes \mathrm{id}_B)(\rho) = \sum_j \lambda_j \Lambda(\rho_j) \otimes \sigma_j \geq 0,$$

since each $\Lambda(\rho_j) \geq 0$.

$\impliedby$: let ρ be entangled. We want to find a positive map $\Lambda : B(\mathbb{H}_A) \to B(\mathbb{H}_A)$ such that $(\Lambda \otimes \mathrm{id}_B)(\rho)$ has a negative eigenvalue. By Definition 2.34, ρ has an entanglement witness ω, with $\mathrm{tr}(\rho\omega) < 0$. By the Choi-Jamiolkowski isomorphism, this defines a map Λ such that

$$\omega = (\Lambda^* \otimes \mathrm{id}_B)(|\phi\rangle\langle\phi|).$$

Since $\mathrm{tr}(XY) = \mathrm{tr}(\mathbb{F}(X \otimes Y))$, and $F = d|\phi\rangle\langle\phi|$, we have for all $A \in B(\mathbb{H}_A)$, $B \in B(\mathbb{H}_B)$,

$$\begin{aligned}
\mathrm{tr}\big(B^T \Lambda(A)\big) &= \mathrm{tr}\big(F\big(\Lambda(A) \otimes B^T\big)\big) \\
&= d\,\mathrm{tr}((\Lambda \otimes \mathrm{id}_B)(A \otimes B)(|\phi\rangle\langle\phi|)) \\
&= d\langle\phi|(\Lambda \otimes \mathrm{id}_B)(A \otimes B)|\phi\rangle.
\end{aligned}$$

TODO: finish. □

Remark 2.37

- In the above proof, we use that $\mathrm{tr}(\rho\omega) = d\langle\phi|(\Lambda \otimes \mathrm{id}_B)(\rho)|\phi\rangle < 0$ implies that $(\Lambda \otimes \mathrm{id}_B)$ has a negative eigenvalue. However, the converse is false. Hence, the positive map Λ corresponding to a witness ω in fact "detects more entanglement" than ω.
- It can be shown that Λ constructed from ω detects an entangled state ρ iff ρ is detected by a witness of the form $(\mathbb{I} \otimes \mathbb{X})\omega(\mathbb{I} \otimes X^*)$ for some $X \in B(\mathbb{H}_B)$.

Remark 2.38 Note that Proposition 2.36 is a theoretical result but is not implementable (in a lab) since Λ is only required to be positive (but not CP). However, the map

$$T(\rho) = \frac{p}{d^2}\mathbb{I}_d \otimes \mathbb{I}_d + (1-p)(\Lambda \otimes \mathrm{id}_B)(\rho)$$

is a CP map. If ρ is separable, then the minimal eigenvalue of $T(\rho)$ must exceed a certain threshold. If it doesn't exceed this threshold, then ρ is entangled.

Remark 2.39 Note that by using a change of abasis via a unitary U, we can obtain a different partial transpose $\tilde{T}_A$ from the "usual" partial transpose T_A:

$$\tilde{T}_A = (U \otimes \mathbb{I})((U^* \otimes \mathbb{I})\rho(U \otimes \mathbb{I}))^{T_A}(U^* \otimes \mathbb{I}) = \left((UU^T) \otimes \mathbb{I}\right)\rho^{T_A}\left(\left(UU^T\right)^* \otimes \mathbb{I}\right) \neq \rho^T$$

Note that this non-uniqueness of the partial transpose does not affect the previous criteria, as they only deal with the eigenvalues, which are invariant under basis changes. Also, we have $\rho^{\tilde{T}_A} \geq 0 \iff \rho^{T_A} \geq 0 \iff \rho^{T_B} \geq 0$, since ρ^{T_A} and ρ^{T_B} differ only by a global transposition.

Definition: Decomposable Map

Definition 2.40 A map $\Lambda : \mathbb{B}(\mathbb{H}) \to \mathbb{B}(\mathbb{H})$ is called **decomposable** if $\Lambda = \Lambda_1 + \Lambda_2 \circ \Theta$, where Λ_1 and Λ_2 are positive maps and Θ is any partial transpose. Otherwise, it is called **non-decomposable**.

Example 2.41 The entanglement witness corresponding to a decomposable map $\Lambda = \Lambda_1 + \Lambda_2 \circ \Theta$ is $\omega = Q_1 + Q_2^T$, where $Q_i = d(\Lambda_i \otimes \mathbb{I})(|\phi\rangle\langle\phi|)$ is the entanglement witness of Λ_i

Proposition: Reduction Criterion

Proposition 2.42 (Reduction Criterion) Let $\Lambda_{\text{red}}(A) = \text{tr}(A)\mathbb{I} - A$. Note that Λ_{red} is positive. Proposition 2.36 gives us

$$(\Lambda_{\text{red}} \otimes \mathbb{I})(\rho) \geq 0 \Longrightarrow \begin{cases} \rho_A \otimes \mathbb{I}_B \geq \rho_{AB} \\ \mathbb{I}_A \otimes \rho_B \geq \rho_{AB}. \end{cases}$$

The entanglement witness corresponding to Λ_{red} is $(\mathbb{I} - F)^{T_A} = 2P_-^{T_A}$, where P_- is the projector onto the anti-symmetric subspace (the space of anti-Hermitian operators). In this case, we obtain

$$\text{tr}(\rho\omega) < 0 \quad \text{iff} \quad \langle\phi|\rho|\phi\rangle \leq \frac{1}{d},$$

where $|\phi\rangle$ is the maximally entangled state.

Proof. Omitted. □

Remark 2.43 If $\mathbb{H} = \mathbb{C}^2 \otimes \mathbb{C}^2$, $P_-^{T_A}$ is 1-dimensional, which gives that entanglement being detected by ω is equivalent to the PPT criterion.

Proposition: Entangled States With Ppt Exist Iff Non Decomposable Maps Exist

Proposition 2.44 Entangled states with positive partial transpose exist iff there are non-decomposable maps. Specifically, there exists a non-decomposable map $T : B(\mathbb{H}_A) \to B(\mathbb{H}_B)$ iff there exists an entangled state $\rho \in B(\mathbb{H}_A) \otimes B(\mathbb{H}_B)$ with positive partial transpose $\rho^{T_A} \geq 0$.

Proof. Omitted. □

Proposition: Separability Criterion In Small Dimensions

Proposition 2.45 Let $\rho \in S(\mathbb{C}^2 \otimes \mathbb{C}^3)$ or $S(\mathbb{C}^2 \otimes \mathbb{C}^2)$. Then ρ is separable iff $\rho^{T_A} \geq 0$.

Proof (Hints). Use the fact that every positive Λ on a Hilbert space of dimension $2 \otimes 2$ or $2 \otimes 3$ is decomposable. $\square$

Proof. This follows from the PPT Criterion and Proposition 2.44 combined with the fact that every positive Λ on a Hilbert space of dimension $2 \otimes 2$ or $2 \otimes 3$ is decomposable. $\square$

3. Quantum hypothesis testing

The goal of quantum hypothesis testing is to distinguish between quantum states by using measurements. Given quantum states, the goal is to minimise the errors in distinguishing them. There are two main frameworks:

- Binary/simple hypothesis testing: we have a null hypothesis ρ_0 and a alternative hypothesis ρ_1. The focus is on minimising either the Type I error (false positive) for a given bound on the Type II error (false negative), or vice versa.
- Quantum state discrimination: states are given with prior probabilities, and the goal is to maximise the probability of correct identification.

3.1. Quantum state discrimination

Given an ensemble $\{\rho_1, ..., \rho_n\} \subseteq \mathbb{S}(\mathbb{H})$ of density operators with corresponding probabilities $\{p_1, ..., p_n\}$, where $p_i \geq 0$ and $\sum_{i=1}^{n} p_i = 1$. This can be interpreted as a set of n hypotheses (the ρ_i) with corresponding a priori probability p_i. The goal is to maximise the average probability of correct identification of the hypothesis. To discriminate among the hypothesis, we use a POVM $M = \{M_1, ..., M_n\}$, and we want to maximise

$$\mathcal{P}(M) := \sum_{j=1}^{n} \mathrm{tr}\big(M_j p_j \rho_j\big) = \sum_{j=1}^{n} p_j \,\mathrm{tr}\big(M_j \rho_j\big).$$

Note that the interpretation is as follows: we have an unknown quantum state ρ which is distributed over $\mathbb{S}(\mathbb{H})$, where $\rho = \rho_i$ with probability p_i. Given that $\rho = \rho_i$, the probability of the measurement M yielding the (correct) outcome i is $\mathrm{tr}\big(M_j \rho_j\big)$. So $\mathcal{P}(M)$ is the expected value of the probability of measuring the correct outcome.

Notation 3.1 Write $\mathcal{M} = \text{span}\Big\{(M_1, ..., M_n) \in \mathbb{B}(\mathbb{H})^n, M_i \geq 0, \sum_i M_i = \mathbb{I}\Big\}$ for the span of the set of POVMs with n operators, and write $\mathcal{P}(\mathcal{M}) = \sup_{M \in \mathcal{M}} \mathcal{P}(M)$.

Notation 3.2 Write $\sigma_i = p_i \rho_i$.

Notation 3.3 For any POVM M, write $L = \sum_{i=1}^{n} M_i p_i \rho_i$, so that $\mathcal{P}(M) = \mathrm{tr}(L)$.

Definition: Optimal Measurement

Definition 3.4 A **maximum likelihood measurement** (or **optimal measurement**) is a measurement (POVM) that achieves the supremum (i.e. the optimal probability) in $\mathcal{P}(\mathcal{M})$.

Proposition: Mle Always Exists

Proposition 3.5 The supremum in $\mathcal{P}(\mathcal{M})$ is always attained, i.e. there is a measurement M^* such that $\mathcal{P}(\mathcal{M}) = \mathcal{P}(M^*)$.

Proof (Hints). Explain why M is compact, the rest is straightforward. □

Proof. For each $M \in \mathcal{M}$, each $M_i \geq 0$, and $\sum_i M_i = \mathbb{I}$, which says that $\mathcal{M}$ is compact. Also, the map $M \mapsto \sum_{i=1}^{n} \mathrm{tr}(M_i p_i \rho_i)$ is linear (and bounded), so is continuous, and so achieves its supremum on $\mathcal{M}$. $\square$

Remark 3.6 Note that since also for each $M \in \mathcal{M}$, each $M_i \geq 0$, we have that $\mathcal{M}$ is convex.

Theorem: Equivalent Conditions For Optimal Measurements

Theorem 3.7 Let $\{\rho_1, ..., \rho_n\}$ be an ensemble with probabilities $\{p_1, ..., p_n\}$. For $M = \{M_1, ..., M_n\}$ and $L = \sum_{i=1}^{n} M_i p_i \rho_i$, TFAE:

1. M is an optimal measurement, i.e. $\mathcal{P}(M) = \mathcal{P}(\mathcal{M})$.
2. For all $i \in [n]$, $\frac{1}{2}(L + L^*) \geq p_i \rho_i$.
3. For all $i \in [n]$, $L \geq p_i \rho_i$.
4. There exists $K \in \mathbb{B}(\mathbb{H})$ such that for all $i \in [n]$, $K \geq p_i \rho_i$ and $(K - p_i \rho_i) M_i = 0$.
5. $\mathcal{P}(M) = \min\{\mathrm{tr}(A) : A \in \mathcal{A}\}$, where $\mathcal{A} = \{A \in \mathbb{B}(\mathbb{H}) : A \geq p_i \rho_i \ \forall i\}$.

Remark 3.8

- The inequalities in 3. and 4. of Theorem 3.7 imply that L and K are Hermitian.
- $L = K$ and are equal to a minimiser in 5. of Theorem 3.7.
- The uniqueness of K does not necessarily imply uniqueness of the optimal measurement.

Proof (Hints). Throughout the proof, use the fact that if $A \leq B$ and $C \geq 0$, then $\operatorname{tr}(AC) \leq \operatorname{tr}(BC)$.

- $1 \Rightarrow 2$: assume the opposite, let P be the orthogonal projector onto the negative eigenspace of $L + L^* - 2p_i\rho_i$. For fixed $\varepsilon > 0$, define $M'_j = (\mathbb{I} - \varepsilon P)M_j(\mathbb{I} - \varepsilon P) + \varepsilon(2 - \varepsilon)P\delta_{ij}$. Verify that M' is a POVM and that

$$\mathcal{P}(M') = \mathcal{P}(M) + \varepsilon \operatorname{tr}(P(2p_i\rho_i - L - L^*)) - \varepsilon^2 \operatorname{tr}(p_i\rho_i P) + \varepsilon^2 \sum_{j=1}^{n} \operatorname{tr}\big(PM_jPp_j\rho_j$$

- $3 \Rightarrow 1$: for any POVM $M' = \{M'_1, ..., M'_n\}$, show that $\mathcal{P}(M) - \mathcal{P}(M') \geq 0$ (recall the properties of a POVM).

- $2 \Rightarrow 1$: use simple modification of the $3 \Rightarrow 1$ proof.
- $2 \Rightarrow 3$: use that

$$\sum_{j=1}^{n} \mathrm{tr}\left(\left(\frac{1}{2}(L + L^*) - p_j \rho_j \right) M_j \right) = \mathrm{tr}\left(\frac{1}{2}(L + L^*) - L \right) = 0$$

  to show that L is Hermitian.
- $3 \Rightarrow 4$: straightforward.
- $4 \Rightarrow 1$: show that $\mathrm{tr}(L) = \mathcal{P}(M)$, show that $\mathcal{P}(M) - \mathcal{P}(M') \geq 0$ for any POVM $M' = \{M'_1, ..., M'_n\}$.
- $4 \Rightarrow 5$: show that $\mathcal{P}(M) = \mathrm{tr}(K)$.
- $5 \Rightarrow 4$: should be straightforward by now.

Proof. Throughout the proof, we use the fact that if $A \leq B$ and $C \geq 0$, then $\mathrm{tr}(AC) \leq \mathrm{tr}(BC)$.

- 1. $\Rightarrow$ 2.: assume the opposite, i.e. that there exists $i \in [n]$ such that $\frac{1}{2}(L + L^*) \not\geq p_i \rho_i$, i.e. $L + L^* - 2p_i\rho_i$ is not positive semi-definite. Let P be the orthogonal projector onto the negative eigenspace of $L + L^* - 2p_i\rho_i$. In particular, P is non-zero. Fix $\varepsilon \in [0, 2]$ and define

$$M'_j = (\mathbb{I} - \varepsilon P) M_j (\mathbb{I} - \varepsilon P) + \varepsilon(2 - \varepsilon) P \delta_{ij}.$$

  It is straightforward to check that M' is a POVM and that

$$\mathcal{P}(M') = \mathcal{P}(M) + \varepsilon \operatorname{tr}(P(2p_i\rho_i - L - L^*)) - \varepsilon^2 \operatorname{tr}(p_i\rho_i P) + \varepsilon^2 \sum_{j=1}^{n} \operatorname{tr}\big(PM_jPp_j\rho_j$$

By construction, $\operatorname{tr}(P(2p_i\rho_i - L - L^*)) \geq 0$. Since the last two terms are $O(\varepsilon^2)$, for ε small enough, $\mathcal{P}(M') > \mathcal{P}(M)$, which contradicts our assumption that $\mathcal{P}(M) = \mathcal{P}(\mathcal{M})$.

- $3 \Rightarrow 1$ and $2 \Rightarrow 1$: let M' be another POVM. Since $\mathcal{P}(M) = \operatorname{tr}(L)$, we have

$$\mathcal{P}(M) - \mathcal{P}(M') = \operatorname{tr}(L) - \sum_{j=1}^{n} \operatorname{tr}\big(M'_j p_j \rho_j\big)$$

$$= \operatorname{tr}\left(L \sum_{j=1}^{n} M'_j \right) - \sum_{j=1}^{n} \operatorname{tr}\big(M'_j p_j \rho_j\big)$$

$$= \sum_{j=1}^{n} \operatorname{tr}\big(M'_j \big(L - p_j \rho_j\big)\big)$$

By 3, $L \geq p_j \rho_j$, hence $\mathcal{P}(M) - \mathcal{P}(M') \geq 0$. For $2 \Rightarrow 1$, since $\operatorname{tr}(L) = \operatorname{tr}(L^*)$, we can replace L in the above proof by $\frac{1}{2}(L + L^*)$.

- $2 \Rightarrow 3$: using that $\operatorname{tr}(L) = \operatorname{tr}(L^*)$, we have

$$\sum_{j=1}^{n} \operatorname{tr}\left(\left(\frac{1}{2}(L + L^*) - p_j \rho_j \right) M_j \right) = \operatorname{tr}\left(\frac{1}{2}(L + L^*) - L \right) = 0$$

Since $\frac{1}{2}(L + L^*) \geq p_j\rho_j$, all the terms $\frac{1}{2}(L + L^*) - p_j\rho_j$ are positive, so $\left(\frac{1}{2}(L + L^*) - p_j\rho_j\right)M_j = 0$ since the sums of the traces are 0. Summing over j gives $\frac{1}{2}(L + L^*) = L$, so L is Hermitian.

- $3 \Rightarrow 4$: choosing $K = L$, it is straightforward to check the conditions are satisfied.
- $4 \Rightarrow 1$: since $KM_j = p_j\rho_j M_j$ for all j, it is straightforward to show that $\mathcal{P}(M) = \mathrm{tr}(L) = \mathrm{tr}(K)$ by summing over j and taking the trace. Letting M' be another POVM, we have

$$\mathcal{P}(M) - \mathcal{P}(M') = \sum_{j=1}^{n} \mathrm{tr}\big(KM'_j\big) - \mathrm{tr}\big(p_j\rho_j M'_j\big)$$

$$= \sum_{j=1}^{n} \operatorname{tr}\big(\big(K - p_j\rho_j\big)M_j'\big) \geq 0$$

since $K - p_j\rho_j \geq 0$.

- $4 \Rightarrow 5$: it is straightforward to show that

$$\mathcal{P}(M) = \operatorname{tr}(K).$$

We have $K \in \mathcal{A}$ and for all $A \in \mathcal{A}$,

$$\operatorname{tr}(K) = \sum_{j=1}^{n} \operatorname{tr}\big(KM_j\big) = \sum_{j=1}^{n} \operatorname{tr}\big(p_j\rho_j M_j\big) \leq \sum_{j=1}^{n} \operatorname{tr}\big(AM_j\big) = \operatorname{tr}(A)$$

So $\mathcal{P}(M) = \operatorname{tr}(K) = \min\{\operatorname{tr}(A) : A \in \mathcal{A}\}$. The argument in reverse shows the converse.

- $5 \Rightarrow 4$: let $A \in \mathcal{A}$ be such that $\operatorname{tr}(A) = \mathcal{P}(M) = \operatorname{tr}(L)$. Then

$$0 = \operatorname{tr}(A - L) = \operatorname{tr}\left(A \sum_{i=1}^{n} M_i - L \right) = \sum_{i=1}^{n} \operatorname{tr}((A - p_i\rho_i)M_i)$$

Since $A \geq p_i\rho_i$ for all i, each term on the RHS is ≥ 0, and so $\operatorname{tr}((A - p_i\rho_i)M_i) = 0$, but $(A - p_i\rho_i)M_i \geq 0$, so we can take $K = A$.

□

Example 3.9 Let $\rho_1, ..., \rho_n$ be pairwise commuting states, so there exists an orthonormal basis $\{|i\rangle : i \in [n]\}$ in which they can be simultaneously diagonalised. Let K be the diagonal operator with diagonal entries $\langle j|K|j\rangle = \max_i \langle j|p_i\rho_i|j\rangle$. By construction, K has minimal trace among all operators A such that $A \geq p_i\rho_i$ for all i (and K is such an operator). Thus, by point 5 of Theorem 3.7,

$$\mathcal{P}(\mathcal{M}) = \min\{\mathrm{tr}(A) : A \geq p_i\rho_i \forall i\} = \mathrm{tr}(K) = \sum_{j=1}^{n} \langle j|K|j\rangle = \sum_j \max_i \langle j|p_i\rho_i|j\rangle.$$

Example 3.10 Let $\rho_1, ..., \rho_n$ be pure states, each with associated a priori probability $1/n$. For simplicity, assume that

$$\sum_{i=1}^{n} p_i \rho_i = \frac{\mathbb{I}_d}{d}$$

(with $d \leq n$). Define $M_i = \frac{d}{n}\rho_i$ for each $i \in [n]$. $\{M_i\}_{i=1}^{n}$ is a POVM which describes a maximum likelihood measurement. Since the ρ_i are pure states, $\rho_i^2 = \rho_i$, so for $L = \sum_{i=1}^{n} M_i p_i \rho_i$, we have

$$L = \sum_{i=1}^{n} M_i p_i \rho_i = \frac{d}{n} \sum_{i=1}^{n} p_i \rho_i^2 = \frac{d}{n} \sum_{i=1}^{n} p_i \rho_i = \frac{\mathbb{I}}{n} \geq p_i \rho_i$$

for all i. Hence, M is an optimal measurement by point 3 of Theorem 3.7.

3.2. Binary hypothesis testing

Let ρ_1 and ρ_2 be density matrices with a priori probability p and $1-p$. Consider the POVM $M = (M_1, M_2) = (P, \mathbb{I} - P)$ with P an orthogonal projection. Assigning P to ρ_1 and $\mathbb{I} - P$ to ρ_2, the probability of error is

$$\mathcal{E}(M) := p \operatorname{tr}(\rho_1(\mathbb{I} - P)) + (1-p) \operatorname{tr}(\rho_2 P).$$

Also,

$$\mathcal{P}(M) = p \operatorname{tr}(\rho_1 P) + (1-p) \operatorname{tr}(\rho_2(\mathbb{I} - P))$$

Note that $\mathcal{P}(M) + \mathcal{E}(M) = 1$.

Definition: Schatten P Norm

Definition 3.11 Let $\mathbb{H}$ be a finite dimensional Hilbert space. For $p \in [1, \infty)$, the **Schatten p-norm** is defined as

$$\|\cdot\|_p : \mathbb{B}(\mathbb{H}) \to [0, \infty),$$

$$\|A\| = \mathrm{tr}(|A|^p)^{1/p}.$$

We can also define $\|A\|_\infty = \lim_{p\to\infty} \|A\|_p = \max_i\{|\lambda_i|\}$, where λ_i are the eigenvalues of A.

Theorem: Quantum Neyman Pearson

Theorem 3.12 (Quantum Neyman-Pearson) We have

$$\mathcal{E}(M) \geq \frac{1}{2}\Big(1 - \|p\rho_1 - (1-p)\rho_2\|_1\Big)$$

with equality iff $P = (p_1\rho_1 - (1-p)\rho_2)_+$ is a projection onto $(p_1\rho_1 - (1-p)\rho_2)_+$, the positive eigensubspace of $p\rho_1 - (1-p)\rho_2$.

Proof (Hints).

- Let $A = p\rho_1 - (1-p)\rho_2$. By considering the positive and negative parts A_+ and A^- of A, show that $\mathrm{tr}(A_+) = \frac{1}{2}(\|A\|_1 + \mathrm{tr}(A))$.
- Also show that $\mathcal{E}(M) = p - \mathrm{tr}(PA)$, and explain why the minimum (over P) of this is attained iff $PA_+ = A_+$ and $PA_- = 0$.

□

Proof. For every Hermitian A, we can write $A = A_+ + A_-$, where A_+ is the positive part and A_- is the negative part. We have

$$\operatorname{tr}(A_+) = \frac{1}{2}(\|A\|_1 + \operatorname{tr}(A))$$

since $\|A\|_1 = \operatorname{tr}(|A|) = \operatorname{tr}(A_+ - A_-)$ and $\operatorname{tr}(A) = \operatorname{tr}(A_+ + A_-)$. Now

$$\begin{aligned}
\mathcal{E}(M) &= p \operatorname{tr}(\rho_1(\mathbb{I} - P)) + (1-p)\operatorname{tr}(p_2 P) \\
&= p - p\operatorname{tr}(\rho_1 P) + (1-p)\operatorname{tr}(p_2 P) \\
&= p - \operatorname{tr}(P(p\rho_1 - (1-p)\rho_2)) =: p - \operatorname{tr}(PA)
\end{aligned}$$

So maximum of above is attained iff $PA_+ = A_+$ and $PA_- = 0$, i.e. P is an orthonormal projection onto A_+. Hence,

$$\begin{aligned}\min_M \mathcal{E}(M) &= p - \mathrm{tr}\Big((p\rho_1 - (1-p)\rho_2)_+\Big) \\ &= p - \frac{1}{2}(\|p\rho_1 - (1-p)\rho_2\|_1 + \mathrm{tr}(p\rho_1 - (1-p)\rho_2) \\ &= \frac{1}{2}\Big(1 - \|p\rho_1 - (1-p)\rho_2\|_1\Big)\end{aligned}$$

Alternatively, we could define $L = Pp\rho_1 + (\mathbb{I} - P)(1-p)\rho_2$ which satisfies $L \geq p\rho_1$ and $L \geq (1-p)\rho_2$, hence is an optimal measurement, hence $1 = \mathcal{P}(M) + \mathcal{E}(M) \leq \mathrm{tr}(L) + \mathcal{E}(M)$. $\square$

Now assume we have m copies of ρ_1 and ρ_2, and we can treat them as single density matrices: $\rho_1^{\otimes m}$ and $\rho_2^{\otimes m}$. For the optimal measurement, the optimal (i.e. minimal) error rate is

$$\mathcal{E}_m^{\text{opt}} = \frac{1}{2}\left(1 - \left\| p\rho_1^{\otimes m} - (1-p)\rho_2^{\otimes m} \right\|_1\right)$$

It can be shown that $\mathcal{E}_m^{\text{opt}}$ decays exponentially with m, i.e. $\mathcal{E}_m^{\text{opt}} \leq Ke^{-\xi m}$, $K, \xi > 0$. Note that this upper bound is independent of p.

Lemma: Quantum Chernoff Bound Lemma

Lemma 3.13 If $A, B \in \mathbb{B}(\mathbb{H})$ are positive, then $\forall s \in [0,1]$, $\mathrm{tr}\big((A^s - B^s)A^{1-s}\big) \leq \mathrm{tr}\big((A-B)_+\big)$.

Proof. Consequence of operator monotonicity of $z \mapsto z^s$ for all $s \in [0, 1]$ (details omitted). $\square$

Theorem: Quantum Chernoff Bound

Theorem 3.14 (Quantum Chernoff Bound) Let $p \neq 0, 1$. Then

$$\xi := \lim_{m\to\infty}\left(-\frac{1}{m}\log(\mathcal{E}_m^{\text{opt}})\right) = -\log\left(\inf_{s\in[0,1]}\operatorname{tr}(\rho_1^{1-s}\rho_2^s)\right) = \max_{s\in[0,1]}(1-s)\overline{D}_s(\rho_1 \parallel \rho_2),$$

where $\overline{D}_s$ is the Petz Renyi divergence.

Proof (Hints).

- Show that $\frac{1}{2}(\mathrm{tr}(A+B) - \|A-B\|_1) \leq \mathrm{tr}\big(B^s A^{1-s}\big)$ for positive $A, B \in \mathbb{B}(\mathbb{H})$ and $s \in [0,1]$.
- Now take $A = p\rho_1^{\otimes m}$ and $B = (1-p)\rho_2^{\otimes m}$ to show inequality in the theorem statement.
- To show equality, consider

$$\hat{\rho}_1 = \sum_{j,k} \lambda_j^{(1)} \left| \langle \psi_j^{(1)} | \psi_k^{(2)} \rangle \right| |jk\rangle\langle jk|$$

$$\hat{\rho}_2 = \sum_{j,k} \lambda_j^{(2)} \left| \langle \psi_j^{(1)} | \psi_k^{(2)} \rangle \right| |jk\rangle\langle jk|,$$

where $\rho_i = \sum_j \lambda_j^{(i)} |\psi_j^{(i)}\rangle\langle\psi_j^{(i)}|$, and use that equality is achieved when applied to commuting operators.

□

Proof. By Lemma 3.13,

$$\begin{aligned}\frac{1}{2}(\operatorname{tr}(A+B)-\|A-B\|_1) &= \frac{1}{2}(2\operatorname{tr}(A)-\operatorname{tr}(A-B)-\operatorname{tr}((A-B)_+)+\operatorname{tr}((A-B)_-\\ &= \operatorname{tr}(A)-\operatorname{tr}((A-B)_+)\\ &\le \operatorname{tr}(A)-\operatorname{tr}((A^s-B^s)A^{1-s}) = \operatorname{tr}(B^sA^{1-s})\end{aligned}$$

Let $A = p\rho_1^{\otimes m}$ and $B = (1-p)\rho_2^{\otimes m}$. Then by above and Quantum Neyman-Pearson,

$$\mathcal{E}_m^{\text{opt}} = \frac{1}{2}\left(1-\left\|p\rho_1^{\otimes m}-(1-p)\rho_2^{\otimes m}\right\|_1\right) \le (1-p)^s p^{1-s}\operatorname{tr}\left(\rho_1^{1-s}\rho_2^s\right)^m$$

Hence

$$\mathcal{E}_m^{\mathrm{opt}} \leq \inf_{s\in[0,1]} p^{1-s}(1-p)^s \mathrm{tr}\big(\rho_1^{1-s}\rho_2^s\big)^m \leq \inf_{s\in[0,1]} \mathrm{tr}\big(\rho_1^{1-s}\rho_2^s\big)^m$$

so

$$-\frac{1}{m}\log \mathcal{E}_m^{\mathrm{opt}} \geq -\log \inf_{s\in[0,1]} \mathrm{tr}\big(\rho_1^{1-s}\rho_2^s\big)$$

And we can take the limit $m \to \infty$.

To show equality: given ρ_1, ρ_2 we can construct $\hat{\rho}_1, \hat{\rho}_2$ such that $[\hat{\rho}_1, \hat{\rho}_2] = 0$ and $\mathrm{tr}(\hat{\rho}_1^{1-s}\hat{\rho}_2^s) = \mathrm{tr}(\rho_1^{1-s}\rho_2^s)$: explicitly, let $\rho_i = \sum_j \lambda_j^{(i)} |\psi_j^{(i)}\rangle\langle\psi_j^{(i)}|$, then we define

$$\hat{\rho}_1 = \sum_{j,k} \lambda_j^1 \left|\langle\psi_j^{(1)}|\psi_k^{(2)}\rangle\right| |jk\rangle\langle jk|$$

$$\hat{\rho}_2 = \sum_{j,k} \lambda_j^2 \left|\langle\psi_j^{(1)}|\psi_k^{(2)}\rangle\right| |jk\rangle\langle jk|,$$

where $\{|ij\rangle\}$ is an orthonormal basis of $\mathbb{H} \otimes \mathbb{H}$. $\hat{\rho}_1, \hat{\rho}_2$ achieve equality in the above inequality. $\square$

3.3. The pretty good measurement

Definition: Pretty Good Measurement

Definition 3.15 Given a collection of states $\{\rho_i\}_{i=1}^n$ with associated prior probability $\{p_i\}_{i=1}^n$, the **pretty good measurement** is $M^P = \{M_i^P\}_{i=1}^n$, where

$$M_i^P = R^{-1/2} p_i \rho_i R^{-1/2} + \frac{1}{n}\left(\mathbb{I} - R^{-1/2} R R^{-1/2}\right) = R^{-1/2} p_i \rho_i R^{-1/2} + \frac{1}{n}\mathbb{I}_{\{\ker R\}}$$

$$R = \sum_{i=1}^n p_i \rho_i,$$

where R^{-1} is the Moore-Penrose pseudo-inverse.

Definition: Square Measurement

Definition 3.16 Given a collection of states $\{\rho_i\}_{i=1}^n$ with associated prior probability $\{p_i\}_{i=1}^n$, the **square measurement** is $M^S = \{M_i^S\}_{i=1}^n$, where

$$M_i^S = S^{-1/2} p_i^2 \rho_i^2 S^{-1/2} + \frac{1}{n}\big(\mathbb{I} - S^{-1/2} S S^{-1/2}\big),$$

$$S = \sum_{i=1}^n p_i^2 \rho_i^2.$$

Theorem: Holders Inequality

Theorem 3.17 (Holder's Inequality) For $p, q \in [1, \infty]$ and $\frac{1}{p} + \frac{1}{q} = 1$, we have

$$\|AB\|_1 = \mathrm{tr}(|AB|) \leq \|A\|_p \|B\|_q.$$

Definition: Operator Convex

Definition 3.18 Let I be an interval. $f : I \to \mathbb{R}$ is **operator convex** on I if

$$f(\lambda A + (1-\lambda)B) \leq \lambda f(A) + (1-\lambda) f(B),$$

for all A, B Hermitian with spectra in I and all $\lambda \in [0,1]$.

Theorem: Jensens Inequality

Theorem 3.19 (Jensen's Inequality) Let f be continuous on an interval I. TFAE:

- f is operator convex on I.
- For each $n \in \mathbb{N}$,

$$f\left(\sum_{i=1}^{n} A_i^* X_i A_i\right) \leq \sum_{i=1}^{n} A_i^* f(X_i) A_i,$$

for all $X_1, ..., X_n$ which are bounded self-adjoint operators whose spectra are contained in I and all operators $A_1, ..., A_n$ are operators which satisfy $\sum_{i=1}^{n} A_i^* A_i = \mathbb{I}$.

- $f(V^*XV) \leq V^*f(X)V$ for all Hermitian X with spectrum in I and all isometries V.

Proposition: Square Measurement Probability Bounds

Proposition 3.20 We have

$$\operatorname{tr}\big(S^{1/2}\big)^2 \leq \mathcal{P}\big(M^S\big) \leq \mathcal{P}_{\text{opt}} \leq \operatorname{tr}\big(S^{1/2}\big).$$

Proof (Hints).

- For simplicity, assume S is invertible. For first inequality, write $S^{1/2} = SS^{-1/2}$, use cyclicity to introduce $\sigma_i^{1/2}$ where appropriate, then use Jensen's Inequality.
- For third inequality, explain why $\sigma_i \leq S^{1/2}$ for each i, and use that for any POVM M, $A \mapsto \mathrm{tr}(M_i A)$ is an operator monotone.

□

Proof. For simplicity, assume S is invertible. The second inequality follows by definition. For the first, we have (letting $\sigma_i = p_i \rho_i$)

$$\begin{aligned}
\operatorname{tr}\left(S^{1/2}\right)^2 = \operatorname{tr}\left(SS^{-1/2}\right)^2 &= \operatorname{tr}\left(\sum_{i=1}^n p_i^2 \rho_i^2 S^{-1/2}\right)^2 \\
&= \left(\sum_{i=1}^n \operatorname{tr}\left(\sigma_i (\sigma_i^{1/2} S^{-1/2} \sigma_i^{1/2})\right)\right)^2 \quad \text{by cyclicity} \\
&\leq \sum_{i=1}^n \operatorname{tr}\left(\sigma_i (\sigma_i^{1/2} S^{-1/2} \sigma_i^{1/2})^2\right) \quad \text{by Jensen's Inequality}
\end{aligned}$$

$$= \sum_{i=1}^{n} \mathrm{tr}\big(\sigma_i^2 S^{-1/2} \sigma_i S^{-1/2}\big) \quad \text{by cyclicity}$$

$$= \sum_{i=1}^{n} \mathrm{tr}\big(\sigma_i M_i^S\big) \quad \text{by cyclicity}$$

$$= \mathcal{P}\big(M^S\big).$$

For the third inequality, note that $\sigma_i^2 \leq \sum_j \sigma_j^2 = S$ for each i, since the σ_i are positive semi-definite. Since $z \mapsto z^{1/2}$ is operator monotone, we have $\sigma_i \leq S^{1/2}$ for each $i \in [n]$. Also, for any POVM $M = \{M_i\}$, $A \mapsto \mathrm{tr}(M_i A)$ is operator monotone, hence $\mathrm{tr}(M_i \sigma_i) \leq \mathrm{tr}\big(M_i S^{1/2}\big)$. Summing over i, we obtain

$$\sum_i \operatorname{tr}(M_i \sigma_i) \leq \sum_i \operatorname{tr}\big(M_i S^{1/2}\big) = \operatorname{tr}\left(\left(\sum_i M_i\right) S^{1/2}\right) = \operatorname{tr}\big(\mathbb{I} \cdot S^{1/2}\big) = \operatorname{tr}\big(S^{1/2}\big).$$

□

Proposition: Pretty Good Measurement Probability Bounds

Proposition 3.21 We have

$$\left(\mathcal{P}_{\text{opt}}\right)^2 \leq \mathcal{P}\left(M^P\right) \leq \mathcal{P}_{\text{opt}}.$$

Proof (Hints). For simplicity, assume R is invertible. For the first inequality, show that for any POVM M, $\left(\sum_{i=1}^{n} \operatorname{tr}(M_i \sigma_i)\right)^2 \leq \mathcal{P}(M^P)$, using cyclicity to introduce $R^{1/4}$ and $R^{-1/4}$ where appropriate, Holder's Inequality, Cauchy-Schwarz, the fact that $\|M_i\|_\infty \leq 1$. Use the fact that $ABA \geq 0$ if $A, B \geq 0$. $\square$

Proof. For simplicity, assume R is invertible. The second inequality follows from the definition. For the first, let $M = \{M_i\}_{i=1}^n$ be a POVM. Then

$$\begin{aligned}
\left(\sum_{i=1}^n \operatorname{tr}(M_i\sigma_i)\right)^2 &= \left(\sum_{i=1}^n \operatorname{tr}\left(\left(R^{1/4}M_iR^{1/4}\right)\cdot\left(R^{-1/4}\sigma_iR^{-1/4}\right)\right)\right)^2 \\
&\leq \left(\sum_{i=1}^n \left\|R^{1/4}M_iR^{1/4}\right\|_2 \left\|R^{-1/4}\sigma_iR^{-1/4}\right\|\right)^2 \quad \text{by Holder} \\
&\leq \sum_{i=1}^n \left\|R^{1/4}M_iR^{1/4}\right\|_2^2 \cdot \sum_{i=1}^n \left\|R^{-1/4}\sigma_iR^{-1/4}\right\|_2^2 \quad \text{by Cauchy-Schwa}
\end{aligned}$$

The first term in the final product is

$$\sum_{i=1}^{n} \left\| R^{1/4} M_i R^{1/4} \right\|_2^2 = \sum_{i=1}^{n} \operatorname{tr}\Big(\left(R^{1/4} M_i R^{1/4} \right)^2 \Big) = \sum_{i=1}^{n} \operatorname{tr}\big(R^{1/2} M_i R^{1/2} M_i \big)$$
$$\leq \sum_{i=1}^{n} \operatorname{tr}\big(R^{1/2} M_i R^{1/2} \big) = \operatorname{tr}(R) = 1,$$

where the inequality follows from Holder's Inequality, since $\|M_i\|_\infty \leq 1$. (Note that $R^{1/4} M_i R^{1/4}$ is PSD since M_i and $R^{1/4}$ are, so can ignore absolute values.) The second term is

$$\sum_{i=1}^{n} \left\| R^{-1/4} \sigma_i R^{-1/4} \right\|_2^2 = \sum_{i=1}^{n} \operatorname{tr}\left(M_i^P \sigma_i \right) = \mathcal{P}(M^P).$$

□

Corollary: Pretty Good And Square Success And Error Bounds

Corollary 3.22 Since $\mathcal{E}(M) = 1 - \mathcal{P}(M)$ and $\mathcal{E}_{\text{opt}} = 1 - \mathcal{P}_{\text{opt}}$, we have

$$\left(P_{\text{opt}}\right)^2 \leq \mathcal{P}(M^P), \mathcal{P}\left(M^S\right) \leq \mathcal{P}_{\text{opt}}, \quad \text{and} \quad \mathcal{E}_{\text{opt}} \leq \mathcal{E}(M^P), \mathcal{E}\left(M^S\right) \leq 2\mathcal{E}_{\text{opt}}.$$

3.4. Asymmetric hypothesis testing

Definition: Type I And Ii Errors

Definition 3.23 Given m copies of states ρ and σ that we want to classify with a POVM $(P_m, \mathbb{I} - P_m)$, the **Type I error** is $\alpha_m(P_m) = \mathrm{tr}(\rho^{\otimes m}(\mathbb{I} - P_m))$, and the **Type II error** is $\beta_m(P_m) = \mathrm{tr}(\sigma^{\otimes m} P_m)$.

Note by the Quantum Chernoff Bound, we have

$$\liminf_{m\to\infty} -\frac{1}{m}\log \alpha_m(P_m) \geq \xi, \quad \liminf_{m\to\infty} -\frac{1}{m}\log \beta_m(P_m) \geq \xi.$$

Theorem: Quantum Steins Lemma

Theorem 3.24 (Quantum Stein's Lemma) Let $\rho, \sigma \in \mathbb{S}(\mathbb{H})$, $\varepsilon \in (0,1)$, let β_m be minimised over all POVMs $(P_m, \mathbb{I} - P_m)$ subject to $\alpha_m(P_m) \leq \varepsilon$. Then

$$\lim_{m\to\infty} -\frac{1}{m} \log \beta_m = D(\rho \parallel \sigma),$$

where $D(\rho \parallel \sigma) = \mathrm{tr}(\rho(\log \rho - \log \sigma))$ is the relative entropy between ρ and σ.

Proof. First we show that $\lim_{m\to\infty} -\frac{1}{m}\log\beta_m \leq D(\rho \parallel \sigma)$.

It can be shown that for A, B positive semi-definite, $\mathrm{tr}\big((A-B)_+\big) \leq \mathrm{tr}\big(A^{1+s}B^{-s}\big)$ for all $s \in [0,1]$. Let $A - B = \sum_i \mu_i Q_i$ be the spectral decomposition of $A - B$, and let $J(X) = \sum_i Q_i X Q_i$ be the pinching on the eigenbasis of $A - B$. This satisfies $[J(A), J(B)] = 0$; also, $\mathrm{tr}\big(A^{1+s}B^{-s}\big)$ is non-increasing under CPTP maps (i.e. $\mathrm{tr}\big(\Phi(A)^{1+s}\Phi(B)^{-s}\big) \leq \mathrm{tr}\big(A^{1+s}B^{-s}\big)$ for all A, B positive semi-definite and quantum channels Φ). We also have $\mathrm{tr}\big((A-B)_+\big) = \mathrm{tr}\big((T(A)-T(B))_+\big)$. Combining these facts, we can assume WLOG that A and B are diagonal matrices. In this case, the inequality

$\mathrm{tr}\big((A-B)_+\big) \le \mathrm{tr}\big(A^{1+s}B^{-s}\big)$ is simply due to the the fact that $a - b \le a(a/b)^s$ for all $a, b > 0$.

Take $A = \rho^{\otimes m}$ and $B = e^{\lambda m}\sigma^{\otimes m}$, with λ a constant to be specified later. Then

$$\begin{aligned}\mathrm{tr}\Big(\big(\rho^{\otimes m} - e^{\lambda m}\sigma^{\otimes m}\big)P_m\Big) &\le \mathrm{tr}\Big((\rho^{\otimes m})^{1+s}e^{-\lambda m s}(\sigma^{\otimes m})^{-s}\Big) \\ &= e^{-\lambda m s}\mathrm{tr}\big(\rho^{1+s}\sigma^{-s}\big)^m\end{aligned}$$

Note that $\alpha_m(P_m) \le \varepsilon$ by assumption, i.e. $1 - \varepsilon \le \mathrm{tr}(\rho^{\otimes m}P_m)$. So by the above inequality,

$$(1-\varepsilon) - e^{\lambda m}\beta_m(P_m) \leq \operatorname{tr}(\rho^{\otimes m}P_m) - e^{\lambda m}\operatorname{tr}(\sigma^{\otimes m}P_m) \leq e^{-\lambda m s}\operatorname{tr}\big(\rho^{1+s}\sigma^{-s}\big)^m$$
$$= e^{-\lambda m s}e^{mf(s)} = e^{m(-\lambda s + f(s))}$$

where $f(s) = \log \operatorname{tr}\big(\rho^{1+s}\sigma^{-s}\big)$. So we have

$$1 - \varepsilon - e^{m(-\lambda s + f(s))} \leq e^{\lambda m}\beta_m(P_m)$$

$$\text{i.e.} \quad \beta_m(P_m) \geq e^{-\lambda m}\Big((1-\varepsilon) - e^{m(f(s)-\lambda s)}\Big)$$

Clearly $f(0) = 0$ and it can be shown that $f'(0) = D(\rho \parallel \sigma)$. So take $\lambda = D(\rho \parallel \sigma) + \delta$ for any $\delta > 0$. Then $\exists s \in (0, 1]$ such that $\lambda s > f(s)$, hence $e^{m(f(s)-\lambda s)} < 1$ for all $m \in \mathbb{N}$. This gives

$$\begin{aligned}\limsup_{m\to\infty} -\frac{1}{m}\log\beta_m(P_m) &\leq \limsup_{m\to\infty} -\frac{1}{m}\log\big(e^{-\lambda m}\big((1-\varepsilon)-e^{m(f(s)-\lambda s)}\big)\big)\\ &= \limsup_{m\to\infty}\bigg(\lambda-\frac{1}{m}\log\big((1-\varepsilon)-e^{m(f(s)-\lambda s)}\big)\bigg)\\ &\leq \lambda \leq D(\rho\parallel\sigma)+\delta.\end{aligned}$$

Since $\delta > 0$ was arbitrary, this shows inequality.

For equality: let $\sigma^{\otimes m} = \sum_{i=1}^{k} \lambda_i P_i$ be the spectral decomposition of $\sigma^{\otimes m}$. Define the completely positive linear map $T : B(\mathbb{H}^{\otimes m}) \to B(\mathbb{H}^{\otimes m})$ by $T(X) = \sum_{i=1}^{k} P_i X P_i$ (this is called a **pinching** on the eigenbasis of $\sigma^{\otimes m}$). Now

$$
\begin{aligned}
D(T(\rho^{\otimes m}) \,\|\, \sigma^{\otimes m}) &= D(T(\rho^{\otimes m}) \,\|\, T(\sigma^{\otimes m})) \leq D(\rho^{\otimes m} \,\|\, \sigma^{\otimes m}) \quad \text{by data-processing} \\
&= mD(\rho \,\|\, \sigma) \quad \text{by addivity} \\
&\leq D(T(\rho^{\otimes m}) \,\|\, \sigma^{\otimes m}) + d\log(m+1).
\end{aligned}
$$

By the inequality, have $D(\rho \,\|\, \sigma) = \lim_{m\to\infty} \frac{1}{m} D(T(\rho^{\otimes m}) \,\|\, \sigma^{\otimes m})$. Also, since the pinching T satisfies $[T(\rho^{\otimes m}), \sigma^{\otimes m}] = 0$, the RHS is interpretable as a classical relative entropy, and classical Stein's lemma has equality. $\square$

4. Distances and entropy measures

Notation 4.1 In this section, log denotes the base two logarithm.

4.1. Quantum entropies

Definition: Von Neumann Entropy

Definition 4.2 Let $\rho \in \mathbb{S}(\mathbb{H})$ be a density matrix. The **von Neumann entropy** of ρ is given by

$$S(\rho) := -\operatorname{tr}(\rho \log \rho).$$

Remark 4.3

- The von Neumann entropy extends the classical Shannon entropy of a discrete probability distribution $\{p_i\}$, given by $-\sum_i p_i \log p_i$. In particular, treating the eigenvalues $\{\lambda_i\}$ of ρ as a distribution, it is easy to see that the von Neumann entropy of ρ is the Shannon entropy of the distribution $\{\lambda_i\}$ (since if $\rho = UDU^*$ for diagonal D and unitary U, then $\log \rho = U \log(D) U^*$, where $\log(D)$ is computed entrywise).
- It quantifies the quantum information content per letter of ensemble (the minimum number of qubits per letter that are necessary to reliably encode a message).

- It also quantifies the amount of entanglement of a bipartite pure state.

Proposition: Properties Of Von Neumann Entropy

Proposition 4.4 Properties of von Neumann entropy: let $\rho \in \mathbb{S}(\mathbb{H})$ with $\dim \mathbb{H} = d$, then

- If $\rho = |\psi\rangle\langle\psi|$ is pure, then $S(\rho) = 0$.
- **Unitary invariance**: $S\big(U\rho U^{-1}\big) = S(\rho)$ for every unitary U.
- $S(\rho) \leq \log d$.
- **Concavity**: for $p_i \geq 0$ and $\sum_i p_i = 1$, we have

$$S\left(\sum_i p_i \rho_i\right) \geq \sum_i p_i S(\rho_i)$$

- **Subadditivity**: for all $\rho_{AB} \in \mathbb{S}(\mathbb{H}_A \otimes \mathbb{H}_B)$, with $\rho_A = \mathrm{tr}_B(\rho_{AB})$ and $\rho_B = \mathrm{tr}_A(\rho_{AB})$,

$$S(\rho_{AB}) \leq S(\rho_A) + S(\rho_B).$$

- **Additivity**: for all $\rho_{AB} = \rho_A \otimes \rho_B \in \mathbb{S}(\mathbb{H}_A \otimes \mathbb{H}_B)$, we have

$$S(\rho_{AB}) = S(\rho_A) + S(\rho_B).$$

- **Strong subadditivity**: for all $\rho_{ABC} \in \mathbb{S}(\mathbb{H}_A \otimes \mathbb{H}_B \otimes \mathbb{H}_C)$,

$$S(\rho_{ABC}) + S(\rho_B) \leq S(\rho_{AB}) + S(\rho_{BC})$$

- **Araki-Lieb inequality**: for all $\rho_{AB} \in \mathbb{S}(\mathbb{H}_A \otimes \mathbb{H}_B)$,

$$S(\rho_{AB}) \geq |S(\rho_A) - S(\rho_B)|.$$

Remark 4.5 If ρ_{AB} is pure, then $S(\rho_{AB}) = 0$ by Proposition 4.4, and so by the Araki-Lieb inequality, $S(\rho_A) = S(\rho_B)$. However, in this case they don't have to be equal to 0.

Corollary: Second Law Of Thermodynamics

Corollary 4.6 Let $\rho_{SE} = \rho_S \otimes \rho_E \in \mathbb{S}(\mathbb{H}_S \otimes \mathbb{H}_E)$. Then by additivity, $S(\rho_{SE}) = S(\rho_S) + S(\rho_E)$. Consider the evolution

$$\rho_{SE} \mapsto U\rho_{SE}U^* = \rho'_{SE}.$$

Then by unitary invariance and subadditivity,

$$S(\rho_S) + S(\rho_E) = S(\rho_{SE}) = S(\rho'_{SE}) \leq S(\rho'_S) + S(\rho'_E).$$

This descrbies the second law of thermodynamics.

Definition: Quantum Mutual Information

Definition 4.7 Let $\rho_{AB} \in \mathbb{S}(\mathbb{H}_A \otimes \mathbb{H}_B)$, $\rho_A = \mathrm{tr}_B(\rho_{AB})$, $\rho_B = \mathrm{tr}_A(\rho_{AB})$. The **quantum mutual information** is

$$I_\rho(A:B) = S(\rho_A) + S(\rho_B) - S(\rho_{AB}).$$

By subadditivity of von Neumann entropy, this is always non-negative.

Definition: Conditional Mutual Information

Definition 4.8 Let $\rho_{ABC} \in \mathbb{S}(\mathbb{H}_A \otimes \mathbb{H}_B \otimes \mathbb{H}_C)$. The **conditional mutual information** is

$$I_\rho(A : C \mid B) := S(\rho_{AB}) + S(\rho_{BC}) - S(\rho_{ABC}) - S(\rho_B)$$

which is always non-negative by strong subadditivity of von Neumann entropy.

Definition: Conditional Entropy

Definition 4.9 For $\rho_{AB} \in \mathbb{S}(\mathbb{H}_A \otimes \mathbb{H}_B)$, the **conditional entropy** is

$$H(A \mid B)_\rho := S(\rho_{AB}) - S(\rho_B).$$

Proposition: Bounds On Conditional Entropy

Proposition 4.10

- $H(A \mid B)_\rho \geq -\log(\dim \mathbb{H}_B)$, so can be negative! This is different to the classical case.
- $S(\rho_A) \geq H(A \mid B)_\rho$.

Definition: Umegaki Relative Entropy

Definition 4.11 The **quantum (Umegaki) relative entropy** of $\rho, \sigma \in \mathbb{S}(\mathbb{H})$ is

$$D(\rho \parallel \sigma) := \begin{cases} \mathrm{tr}(\rho(\log \rho - \log \sigma)) & \text{if} \ker \sigma \subseteq \ker \rho \\ \infty & \text{otherwise.} \end{cases}$$

It is a quantum analogue of the classical relative entropy (the Kullback-Leibler divergence).

Definition: Belavkin Staszewski Relative Entropy

Definition 4.12 The **Belavkin-Staszewski relative entropy** of $\rho, \sigma \in \mathbb{S}(\mathbb{H})$ is

$$\hat{D}(\rho \parallel \sigma) = \mathrm{tr}\big(\rho \log\big(\rho^{1/2}\sigma^{-1}\rho^{1/2}\big)\big).$$

Proposition: Umegaki Entropy Upper Bound

Proposition 4.13 We have $D(\rho \parallel \sigma) \leq \widehat{D}(\rho \parallel \sigma)$, with equality iff $[\rho, \sigma] = 0$.

Proposition: Properties Of Quantum Relative Entropy

Proposition 4.14 Properties of quantum relative entropy:

- $\rho \mapsto D(\rho \parallel \sigma)$ and $\rho \mapsto \hat{D}(\rho \parallel \sigma)$ are continuous.
- **Non-negativity**: $D(\rho \parallel \sigma) \geq 0$ with equality iff $\rho = \sigma$, and similarly for $\hat{D}$.
- **Unitary invariance**: $D(U\rho U^* \parallel U\sigma U^*) = D(\rho \parallel \sigma)$, and similarly for $\hat{D}$.
- **Additivity**:

$$D(\rho_A \otimes \rho_B \parallel \sigma_A \otimes \sigma_B) = D(\rho_A \parallel \sigma_A) + D(\rho_B \parallel \sigma_B),$$

  and similarly for $\hat{D}$.
- **Superadditivity**:

$$D(\rho_{AB} \parallel \sigma_A \otimes \sigma_B) \geq D(\rho_A \parallel \sigma_A) + D(\rho_B \parallel \sigma_B).$$

- **Data-processing inequality**: for every quantum channel T and $\rho, \sigma \in \mathbb{S}(\mathbb{H})$,

$$D(T(\rho) \parallel T(\sigma)) \leq D(\rho \parallel \sigma),$$

and similarly for $\hat{D}$.

Remark 4.15 The quantum (Umegaki) relative entropy can be defined as the unique (up to normalisation) function satisfying the properties of Proposition 4.14.

4.2. Divergences

Definition: RéNyi Divergence

Definition 4.16 For $\alpha \in (0,1) \cup (1,\infty)$, the **Rényi divergence** between PMFs $\{p_i\}$ and $\{q_i\}$ is

$$D_\alpha(\{p_i\} \parallel \{q_i\}) = \frac{1}{\alpha - 1} \log \frac{\sum_i p_i^\alpha q_i^{1-\alpha}}{\sum_i p_i}.$$

Note that $\lim_{\alpha \to 1} D_\alpha(p \parallel q) = D(p \parallel q)$ where

$$D(p \parallel q) = \sum_i p_i \log \frac{p_i}{q_i}$$

is the Kullback-Leibler divergence.

Definition: Axioms Of Quantum Renyi Divergence

Definition 4.17 A **quantum divergence** (with parameter α) is a function $\mathbb{D}_\alpha(\cdot \,\|\, \cdot) : \mathbb{S}(\mathbb{H}) \times \mathbb{S}(\mathbb{H}) \to \mathbb{R}_{\geq 0}$ which can be expressed as $\mathbb{D}_\alpha(\cdot \,\|\, \cdot) = g(Q(\cdot \,\|\, \cdot))$ where g is continuous and strictly monotone, and satisfies:

- **Positive semidefiniteness**: $\mathbb{D}_\alpha(\rho \,\|\, \sigma) \geq 0$ with equality iff $\rho = \sigma$.
- **Data processing inequality**: $\mathbb{D}_\alpha(T(\rho) \,\|\, T(\sigma)) \leq \mathbb{D}_\alpha(\rho \,\|\, \sigma)$ for every quantum channel T.
- **Joint convexity**: if $\alpha > 1$, then for all $\{p_i\}$ such that each $p_i \geq 0$ and $\sum_i p_i = 1$,

$$Q\left(\sum_i p_i\rho_i \parallel \sum_i p_i\sigma_i\right) \leq \sum_i p_i Q(\rho_i \parallel \sigma_i),$$

- **Joint concavity**: if $\alpha \in (0,1)$, then for all $\{p_i\}$ such that each $p_i \geq 0$ and $\sum_i p_i = 1$,

$$Q\left(\sum_i p_i\rho_i \parallel \sum_i p_i\sigma_i\right) \geq \sum_i p_i Q(\rho_i \parallel \sigma_i),$$

- **Dominance**: for all $X, Y, Y' \in \mathbb{B}(\mathbb{H})$ positive with $Y \leq Y'$, we have $\mathbb{D}_\alpha(X \parallel Y) \geq \mathbb{D}_\alpha(X \parallel Y')$.

Note that unlike in the classical case, there are multiple families of functions which satisfy these axioms.

Definition: Pinching Map

Definition 4.18 Given orthogonal projections $P_1, ..., P_n$ (i.e. each $P_i = P_i^*$ and $\sum_i P_i = \mathbb{I}$), we define the **pinching map** $\mathcal{P} : \mathbb{B}(\mathbb{H}) \to \mathbb{B}(\mathbb{H})$ by

$$\mathcal{P}(L) = \sum_{j=1}^{n} P_j L P_j = \sum_{k=1}^{n} U_k L U_k^*,$$

where $U_k = \sum_{j=1}^{n} e^{2\pi i j k/n} P_j$. It is easy to see that $\mathcal{P}$ is a CPTP map.

Remark 4.19 For $H = \sum_j \lambda_j |e_j\rangle\langle e_j|$ Hermitian, we can set $P_\lambda = \sum_{j:\lambda_j=\lambda} |e_j\rangle\langle e_j|$ so that $H = \sum_\lambda \lambda P_\lambda$. We can then define the pinching map $\mathcal{P}_H : L \to \sum_\lambda P_\lambda L P_\lambda$ of H. This has the following properties:

- $P_H(H) = H$.
- $P_H(L) \geq \frac{1}{|\mathrm{spec}(H)|} L$: note that $|\mathrm{spec}(H^{\otimes n})| = O(\mathrm{poly}(n))$.
- $[P_H(L), H] = 0$.
- $\mathrm{tr}(P_H(L)H) = \mathrm{tr}(LH)$.
- We can express P_H as an integral: $P_H(L) = \int H^{it} L H^{-it} \mu(\mathrm{d}t)$.

Definition: Preparation Map

Definition 4.20 For $\rho, \sigma \in \mathbb{S}(\mathbb{H})$, let $\Delta = \sigma^{-1/2}\rho\sigma^{-1/2} = \sum_{k=1}^{m} \lambda_k \Pi_k$ be the spectral decomposition. Define $q_k = \mathrm{tr}(\sigma\Pi_k)$ and $p_k = \lambda_k q_k$ for each $k \in [m]$ (note p and q can be interpreted as classical probability distributions). The **preparation map** of ρ and σ is

$$\Lambda(L) = \sum_{k=1}^{m} \langle k|L|k\rangle \frac{1}{q(k)} \sigma^{1/2}\Pi_k\sigma^{1/2}.$$

It is easy to see that $\Lambda(p) = \rho)$ and $\Lambda(q) = \sigma$ (where we interpret p and q as the matrices $\mathrm{diag}(p) = \sum_{k=1}^{m} p_k|k\rangle\langle k|$ and $\mathrm{diag}(q) = \sum_{k=1}^{m} q_k|k\rangle\langle k|$). Note that Λ is a CPTP map.

Definition: Minimal RéNyi Divergence

Definition 4.21 For $\rho, \sigma \in \mathbb{S}(\mathbb{H})$ and $\alpha \in (0,1) \cup (1,\infty)$, the **minimal/sandwiched Rényi divergence** is

$$\rho \parallel \sigma) := \frac{1}{\alpha - 1} \log \operatorname{tr}\Big(\big(\sigma^{(1-\alpha)/2\alpha} \rho \sigma^{(1-\alpha)/2\alpha} \big)^{\alpha} \Big) = \lim_{n\to\infty} \frac{1}{n} \mathbb{D}_{\alpha}(\mathcal{P}_{\sigma^{\otimes n}}(\rho^{\otimes n}) \parallel \sigma^{\otimes n})$$

$$= \lim_{n\to\infty} \frac{1}{n} \mathbb{D}_{\alpha}(\mathcal{P}_{\sigma^{\otimes n}}(\rho^{\otimes n}) \parallel \mathcal{P}_{\sigma^{\otimes n}}(\sigma^{\otimes}$$

where $\mathcal{P}_{\sigma^{\otimes n}}$ is the pinching map of $\sigma^{\otimes n}$ and $\mathbb{D}_{\alpha}$ is *any* quantum divergence. By the data processing inequality, this is at most $\lim_{n\to\infty} \frac{1}{n} \mathbb{D}_{\alpha}(\rho^{\otimes n} \parallel \sigma^{\otimes n}) = \mathbb{D}_{\alpha}(\rho \parallel \sigma)$.

Definition 4.22 For $\alpha \in (1,2)$, the **geometric mean** of $\rho, \sigma \in \mathbb{S}(\mathbb{H})$ is

$$\sigma \#_\alpha \rho = \sigma^{1/2} \left(\sigma^{-1/2} \rho \sigma^{-1/2} \right)^\alpha \sigma^{1/2}.$$

Definition: Maximal Renyi Divergence

Definition 4.23 For $\rho, \sigma \in \mathbb{S}(\mathbb{H})$ and $\alpha \in (1,2)$, the **maximal/geometric Rényi divergence** of ρ and σ is

$$\hat{D}_\alpha(\rho \parallel \sigma) = \frac{1}{\alpha - 1} \log \operatorname{tr}\Big(\sigma\big(\sigma^{-1/2}\rho\sigma^{-1/2}\big)^\alpha\Big) = \frac{1}{\alpha - 1} \log \operatorname{tr}(\sigma \#_\alpha \rho) = \mathbb{D}_\alpha(p \parallel q),$$

where p and q are as in Definition 4.20 and $\mathbb{D}_\alpha$ is *any* quantum divergence. (Again, we interpret p and q as diagonal matrices.) Note that this is at least $\mathbb{D}_\alpha(\rho \parallel \sigma) = \mathbb{D}_\alpha(\Lambda(p) \parallel \Lambda(q))$ by the (classical) data processing inequality, where Λ is the preparation map of ρ and σ.

Proposition: Properties Of Maximal And Minimal Renyi Divergence

Proposition 4.24

- We have $\hat{D}_\alpha(\rho \parallel \sigma) \geq \tilde{D}_\alpha(\rho \parallel \sigma)$ with equality iff $[\rho, \sigma] = 0$.
- $\lim_{\alpha \to 1} \hat{D}_\alpha(\rho \parallel \sigma) = \hat{D}(\rho \parallel \sigma)$.
- $\lim_{\alpha \to 1} \tilde{D}_\alpha(\rho \parallel \sigma) = D(\rho \parallel \sigma)$.

Definition: Petz RéNyi Divergence

Definition 4.25 For $\alpha \in (0,1) \cup (1,\infty)$ and $\rho, \sigma \in \mathbb{S}(\mathbb{H})$, the **Petz Rényi divergence** is

$$\overline{D}_\alpha(\rho \parallel \sigma) = \frac{1}{\alpha - 1} \log \operatorname{tr}\left(\rho^\alpha \sigma^{1-\alpha}\right).$$

We have $\lim_{\alpha \to 1} \overline{D}_\alpha(\rho \parallel \sigma) = D(\rho \parallel \sigma)$.

Definition: Max Relative Entropy

Definition 4.26 The **max relative entropy** is

$$D_{\max}(\rho \parallel \sigma) = \inf\{\lambda > 0 : \rho \leq 2^{\lambda}\sigma\}.$$

We have $D_{\max}(\rho \parallel \sigma) = \lim_{\alpha \to \infty} \overline{D}_{\alpha}(\rho \parallel \sigma)$.

Definition: Quantum Divergences Ordering

Proposition 4.27 For all $\alpha \in (1,2)$,

$$D(\rho \parallel \sigma) \leq \tilde{D}_\alpha(\rho \parallel \sigma) \leq \overline{D}_\alpha(\rho \parallel \sigma) \leq \hat{D}_\alpha(\rho \parallel \sigma) \leq D_{\max}(\rho \parallel \sigma).$$

4.3. Applications

Recall the setting of binary hypothesis testing from the previous chapter. Given n copies of the states $\rho, \sigma \in \mathbb{S}(\mathbb{H})$, the null hypothesis is $\rho^{\otimes n}$ and the alternative hypothesis is $\sigma^{\otimes n}$. We want to discriminate between both by considering POVMs $M_n = \{P_n, \mathbb{I} - P_n\}$, with P_n an orthogonal projection.

Type I (first kind) error occurs if we wrongly conclude that alternative hypothesis is correct, with probability $\alpha_n(P_n; \rho) := \mathrm{tr}(\rho^{\otimes n}(\mathbb{I} - P_n))$. Type II (second kind) error occurs if we wrongly conclude that the null hypothesis is correct, with probability $\beta_n(P_n; \sigma) := \mathrm{tr}(\sigma^{\otimes n} P_n)$.

By Quantum Neyman-Pearson, assuming that the a priori probabilities of ρ and σ are both $1/2$, we have

$$\min_{M_n \text{ POVM}} \frac{1}{2}(\alpha_n(P_n;\rho) + \beta_n(P_n;\sigma)) = \frac{1}{2}\Big(1 - \|p\rho^{\otimes n} - (1-p)\sigma^{\otimes n}\|_1\Big),$$

and taking the limit as $n \to \infty$ gives, by the Quantum Chernoff Bound,

$$\begin{aligned}\lim_{n\to\infty} -\frac{1}{n}\log \min_{M_n \text{ POVM}} \frac{1}{2}(\alpha_n(P_n;\rho) + \beta_n(P_n;\sigma)) &= \max_{s\in[0,1]} -\log \operatorname{tr}\left(\rho^s \sigma^{1-s}\right) \\ &= -\min_{s\in[0,1]} \log \overline{Q}_s(\rho \parallel \sigma)\end{aligned}$$

$$= \max_{s \in [0,1]} (1-s) \overline{D}_s(\rho \parallel \sigma),$$

where $\overline{D}_s(\rho \parallel \sigma)$ is the Petz Rényi divergence, and $\overline{Q}_s(\rho \parallel \sigma) = \operatorname{tr}\left(\rho^s \sigma^{1-s}\right)$ is as in Definition 4.17.

So the interpretation of the Petz Rényi divergence is that it provides optimal exponential rate for the combined errors in binary hypothesis testing.

Similarly, we can consider the setting of asymmetric hypothesis testing. By the Quantum Stein's Lemma,

$$\lim_{n\to\infty} -\frac{1}{n}\log\beta_n(P_n;\sigma) = D(\rho \parallel \sigma),$$

so the relative entropy can be interpreted as being the optimal exponential rate for the type II error in binary hypothesis, subject to the type I error being bounded by a certain constant.

For probability distributions P and Q and $\alpha \in (1,\infty)$, we have $D_\alpha(P \parallel Q) = \frac{1}{\alpha-1}\sum_x P(x)\Big(\frac{P(x)}{Q(x)}\Big)^{\alpha-1}$. We have the chain rule for classical Rényi divergence:

$$D_\alpha\big(P_{XY} \parallel Q_{X,Y}\big) \le D_\alpha(P_X \parallel Q_X) + \max_x D\Big(P_{Y\,|\,X=x} \parallel Q_{Y\,|\,X=x}\Big).$$

Definition: Quantum Channel Divergence

Definition 4.28 Let $\mathbb{D}_\alpha$ be a quantum Rényi divergence. We can extend this to a **Rényi divergence of quantum channels** by defining, for quantum channels T_1, T_2,

$$\mathbb{D}_\alpha(T_1 \parallel T_2) := \sup_{\rho \in \mathbb{S}(\mathbb{H})} \mathbb{D}_\alpha(T_1(\rho) \parallel T_2(\rho)).$$

Lemma: Matsumoto

Lemma 4.29 (Matsumoto) For $\rho, \sigma \in \mathbb{S}(\mathbb{H})$, $\alpha \in (0,1) \cup (1,2]$, we have

$$\hat{D}_\alpha(\rho \parallel \sigma) = \inf_{P,Q,\Gamma} D_\alpha(P \parallel Q)$$

where $\Gamma(P) = \rho$, $\Gamma(Q) = \sigma$, Γ is a quantum channel.

The infimum is attained by the preparation map Λ of ρ and σ: let $\sigma^{-1/2}\rho\sigma^{-1/2} = \sum_x \lambda_x \Pi_x$, define $\Lambda(L) = \sum_x \frac{\langle x|L|x\rangle}{Q(x)} \sigma^{1/2}\Pi_x\sigma^{1/2}$, with $Q(x) = \mathrm{tr}(\sigma\Pi_x)$, $P(x) = \lambda_x Q(x)$.

Theorem: Chain Rule For Quantum Channels

Theorem 4.30 (Chain Rule for Quantum Channels) Let $\rho, \sigma \in \mathbb{S}(\mathbb{H})$, T_1, T_2 be quantum channels and $\alpha \in (0,1) \cup (1,\infty)$, then

$$\mathbb{D}_\alpha(T_1(\rho) \parallel T_2(\sigma)) \leq \hat{D}_\alpha(\rho \parallel \sigma) + \mathbb{D}_\alpha(T_1 \parallel T_2).$$

for any quantum Rényi divergence $\mathbb{D}_\alpha$. In particular, $\hat{D}_\alpha(T_1(\rho) \parallel T_2(\sigma)) \leq \hat{D}_\alpha(\rho \parallel \sigma) + \hat{D}_\alpha(T_1 \parallel T_2)$.

Proof (Hints).

- Let P, Q, Λ be as in Matsumoto, show that $T_1(\rho) = \sum_x P(x) T_1(\Lambda(|x\rangle\langle x|))$, similarly for $T_2(\sigma)$.
- Explain why

$$\mathbb{D}_\alpha(T_1(\rho) \parallel T_2(\sigma)) \leq \frac{1}{\alpha - 1} \log \sum_x P(x)^\alpha Q(x)^{1-\alpha} \cdot 2^{\mathbb{D}_\alpha(T_1(\Lambda(|x\rangle\langle x|)) \parallel T_2(\Lambda(|x\rangle\langle x|)))},$$

split the sum by the taking a maximum over the second terms in the sum.

□

Proof. By Matsumoto, there are P and Q such that $\Lambda(P) = \rho$ and $\Lambda(Q) = \sigma$. So we have, by linearity of quantum channels,

$$T_1(\rho) = T_1(\Lambda(P)) = T_1\left(\Lambda\left(\sum_x P(x)|x\rangle\langle x|\right)\right) = \sum_x P(x) T_1(\Lambda(|x\rangle\langle x|))$$

$$T_2(\sigma) = T_2(\Lambda(Q)) = T_2\left(\Lambda\left(\sum_x Q(x)|x\rangle\langle x|\right)\right) = \sum_x Q(x) T_2(\Lambda(|x\rangle\langle x|))$$

Then (since we can treat these classically as they are diagonal)

$$\begin{aligned}
\mathbb{D}_\alpha(T_1(\rho) \parallel T_2(\sigma)) &= \mathbb{D}_\alpha\left(\sum_x P(x)T_1(\Lambda(|x\rangle\langle x|)) \middle\| \sum_x Q(x)T_2(\Lambda(|x\rangle\langle x|))\right) \\
&\leq \frac{1}{\alpha-1}\log\sum_x P(x)^\alpha Q(x)^{1-\alpha} \cdot 2^{\mathbb{D}_\alpha(T_1(\Lambda(|x\rangle\langle x|)) \parallel T_2(\Lambda(|x\rangle\langle x|)))} \\
&\leq D_\alpha(P \parallel Q) + \max_x \mathbb{D}_\alpha(T_1(\Lambda(|x\rangle\langle x|)) \parallel T_2(\Lambda(|x\rangle\langle x|))) \\
&\leq \hat{D}_\alpha(\Lambda(\rho) \parallel \Lambda(\sigma)) + \mathbb{D}_\alpha(T_1 \parallel T_2) \\
&\leq \hat{D}_\alpha(\rho \parallel \sigma) + \mathbb{D}_\alpha(T_1 \parallel T_2),
\end{aligned}$$

where the final inequality is by data-processing. □

Definition: Regularised Renyi Divergence

Definition 4.31 For a quantum Renyi divergence $\mathbb{D}_\alpha$, the **regularised Rényi divergence** between quantum channels T_1 and T_2 is

$$\mathbb{D}_\alpha^{\mathrm{reg}}(T_1 \parallel T_2) := \lim_{n\to\infty} \frac{1}{n}\mathbb{D}_\alpha(T_1^{\otimes n} \parallel T_2^{\otimes n})$$

This equals $\mathbb{D}_\alpha(T_1 \parallel T_2)$ if $\mathbb{D}_\alpha$ is the geometric Rényi divergence, but this generally does not hold for other quantum divergences.

Theorem: Schumacher

Theorem 4.32 (Schumacher) Let $\{|\psi(x)\rangle, p(x)\}_x$ be an ensemble of states $|\psi(x)\rangle$ with probabilities $p(x)$. Let $\rho = \sum_x p(x)|\psi(x)\rangle\langle\psi(x)|$. Then the optimal rate of compression is given by the von Neumann entropy of ρ.