1 Measurement

1.1 General Measurements

A general measurement is described by a collection of measurement operators $\{M_m\}$ such that

$$p(m) = \langle\psi| M_m^\dagger M_m |\psi\rangle$$

The state after measurement is

$$\frac{M_m |\psi\rangle}{\sqrt{p(m)}}$$

And the operators satisfy a completeness

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

1.2 Projective Measurement

A projective measurement is a special case of the above, in which the masurement can be interpreted as measuring the value of some observable quantity (of only the system in question).

An observable (Hermitian operator) H can be decomposed $H = \sum_m h_m P_m$, where the P_m are projectors and the h_m are the relevant eigenvalues. Measurement of the observable can be viewed as a general measurement in which $M_m = P_m$, that is the M_m satisfy $M_i M_j = \delta_{ij} M_j$. This allows one to easily calculate observable expectation values:

$$E(H) = \langle\psi| H |\psi\rangle$$

The ability to perform projective measurements, when combined with the ability to enlarge the space and perform unitaries, enables one to perform projective measurements.

1.3 POVM

Positive-operator valued measurements is a formalism for when we don't actually care about the state after measurement. We just compress the $M_m^\dagger M_m$ into the positive operators E_m, and say a POVM is any set of positive $\{E_m\}$ such that they sum to the identity.

So $p(m) = \langle\psi| E_m |\psi\rangle$ and $\sum_m E_m = I$. Iff the measurement is a projective measurement, the E_m satisfy $E_i E_j = \delta_{ij} E_j$ (because the E_m are just the P_m themselves).

2 Universal quantum gates

2.1 Errors

$$E(U, V) = \max_{|\psi\rangle} ||(U - V) |\psi\rangle ||$$

Under this definition, the difference in the probability of any measurement outcome is less than 2E, and errors add, at worst, linearly as operations are compounded.

2.2 Acheiving universality to arbitrary precision

Hadamard, CNOT, and $\pi/8$ are universal.

First, an arbitary unitary on n qubits can be expressed *exactly* as a product of k two-level unitaries ($k \leq d(d-1)/2$). A two-level unitary operates non-trivially on only two basis states. *Note: since $d = 2^n$, this step may require exponentially many gates in general.*

Second, an arbitrary two-level unitary can be written as a sequence of single-qubit unitaries and CNOT gates.

Finally, an arbitrary single-qubit unitary can be arbitrarily well-approximated by the Hadamard gate and the $\pi/8$ gate. T is a $\pi/4$ rotation about $\hat{z}$ and $HTHT$ is a $\pi/4$ rotation about $\hat{x}$. Combining them yields a certain irrational rotation about a certain axis, which can be repeated to approximate any rotation about that axis. $H(HTHT)H$ is an irrational rotation about an orthogonal axis, and we can use the result that any single-qubit unitary can be written as a product $R_{\hat{n}}R_{\hat{m}}R_{\hat{n}}$ to approximate any unitary.

The Solovay-Kitaev theorem states that any single-qubit unitary can be approximated with accuracy ε in $O(\log^c(1/\varepsilon))$ gates. Since errors add, at worst, linearly, approximating an entire m-gate circuit to ε requires $O(m\log^c(m/e))$ operations. Note: m may, in general, depend exponentially on n, so arbitrary n-qubit unitaries are difficult to approximate.

3 Fourier Transform Algorithms

3.1 Quantum Fourier Transform

The quantum fourier transform is given by

$$|j\rangle \to \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} e^{2\pi ijk/N} |k\rangle$$

or, equivalently

$$\sum_{j=0}^{N-1} x_j |j\rangle \to \sum_{k=0}^{N-1} y_k |k\rangle$$

where

$$y_k = \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} x_j e^{2\pi ijk/N}$$

is the classical Fourier transform. It is often convenient to use the product representation:

$$\begin{aligned} |j_1, j_2 \cdots, j_n\rangle \to & \quad \frac{1}{2^{n/2}} \bigotimes_{l=1}^{n} \left(|0\rangle + e^{2\pi ij2^{-l}} |1\rangle\right) \\ = & \quad \frac{\left(|0\rangle + e^{2\pi i 0.j_n} |1\rangle\right)\left(|0\rangle + e^{2\pi i 0.j_{n-1}j_n} |1\rangle\right) \cdots \left(|0\rangle + e^{2\pi i 0.j_1 j_2 \ldots j_n} |1\rangle\right)}{2^{n/2}} \end{aligned}$$

3.2 Phase estimation

The phase estimation procedure estimates the eigenvalue $e^{i2\pi\varphi}$ corresponding to an eigenvector $|u\rangle$ of U. An ancilliary set of qubits is Hadamarded and the j-th qubit controls an operation U^{2^j} upon $|u\rangle$, and the phase kicks back onto that j-th qubit, so at the end of the operation, the ancilliary bits contain

$$\frac{1}{2^{t/2}} \sum_{k=0}^{2^t-1} e^{2\pi i\varphi k} |k\rangle$$

after which, an inverse Fourier transform should give φ in the computational basis.

For φ which are not expressible as a t-qubit fraction, this algorithm returns a close estimate to φ‘. Explicitly, if we want φ to n bits with probability of success $1-\varepsilon$, then we want $t = n + \lceil \log(2 + 1/2\varepsilon) \rceil$ auxilary qubits. The runtime will then be $O(t^2)$.

3.3 Order-finding

The order of x modulo N (where x and N are coprime) is the smallest integer r such that $x^r = 1 \pmod N$. Order-finding can be accomplished via phase estimation upon the following unitary:

$$U|y\rangle = |xy \pmod N\rangle$$

For $N > y$, I otherwise.

(Since $U^r|y\rangle = |x^r y \pmod N\rangle = |y\rangle$, $U^r = 1$, so the eigenvalues are r-th roots of unity). Eigenvectors of U are

$$|u_s\rangle = \frac{1}{\sqrt{r}} \sum_{k=0}^{r-1} \exp\left\{\frac{-2\pi isk}{r}\right\} |x^k \pmod N\rangle \tag{1}$$

with eigenvalues

$$\exp\left\{\frac{2\pi is}{r}\right\}$$

from which one can determine r.

To execute the algorithm, one needs (1) an efficient way to execute U^{2^j},modular exponentiation; and, more difficultly, (2) a way to generate the state $|u_s\rangle$, which would require knowing r! Fortunately, Eq. 1 implies that

$$\frac{1}{\sqrt{r}} \sum_{s=0}^{r-1} |u_s\rangle = |1\rangle$$

so preparing $|1\rangle$ should give a superposition of estimates φ of s/r. From there, a continued-fractions algorithm can determine r, with various tricks to handle that s and r may not always be coprime.

Factoring numbers can be reduced, via a bit of number theory, to an order-finding problem, and voila, RSA is broken.

3.4 Period-finding

Let $f(x)$ be a function with single-bit output such that $f(x+r) = f(x)$.

Begin in $|0\rangle|0\rangle$, and create the superposition:

$$\frac{1}{\sqrt{2^t}} \sum_{x=0}^{2^t-1} |x\rangle |0\rangle$$

Apply U to get

$$\frac{1}{\sqrt{2^t}} \sum_{x=0}^{2^t-1} |x\rangle |f(x)\rangle$$

Now, since f(x) is periodic, the $|f(x)\rangle$ factors group the $|x\rangle$ factors into groups of $\{x, x+r, x+2r\cdots\}$, so that applying an (inverse) Fourier transform to the x register will set it into a superposition of computational basis states representing binary fractions corresponding to integer multiples of $1/r$ (with amplitudes modulated by the higher Fourier components of $f(x)$).

From there, a continued fractions algorithm will calculate r.

4 Searching Algorithms

For searching, we will have an "oracle" U such that $U|x\rangle|y\rangle = |x\rangle|y + f(x)\rangle$ where $f(x)$ is the binary answer to "is this a solution?" We will always have $|y\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$, so, using the phase-kickback technique, we will ignore this space and just write the action of the oracle as $U|x\rangle = (-1)^{f(x)}|x\rangle$

4.1 Grover Search

Apply Hadamard to produce an equal superposition $|\varphi\rangle = \frac{1}{\sqrt{N}} \sum_x |x\rangle$. Now, if there are M solutions in the space, and $N-M$ non-solutions, then we can write $|\varphi\rangle$ as a superposition of the equally weighted non-solution vector $|\alpha\rangle$ and the equally weighted solution vector $|\beta\rangle$. In terms of these quantities

$$|\varphi\rangle = \sqrt{\frac{N-M}{N}}|\alpha\rangle + \sqrt{\frac{M}{N}}|\beta\rangle$$

The Grover iteration (G) proceeds as follows: First, apply the oracle. This negates the coefficient on $|\beta\rangle$, effecting a reflection over $|\alpha\rangle$. Then apply the operator $H^{\otimes n}(I - 2|0\rangle\langle 0|)H^{\otimes n}$, which effects a reflection over $|\varphi\rangle$.

The result is a rotation in the α–β plane by θ where $\cos\theta/2 = \sqrt{(N-M)/N}$. Since the initial state is $|\varphi\rangle$, which can be written $\cos\theta/2|\alpha\rangle + \sin\theta/2|\beta\rangle$, the result of applying G k times is $\cos(k+1/2)\theta|\alpha\rangle + \sin(k+1/2)\theta|\beta\rangle$. We would like to rotate the state to $|\beta\rangle$.

Taylor expanding for $M \ll N$, we find $\theta \approx 2\sqrt{M/N}$. So we rotate $\sim (\pi/2)/\theta = \pi/4\sqrt{N/M}$ times. And we have an angular error of $\theta/2 = \sqrt{M/N}$ which gives an error probability of M/N.

We assumed $M \ll N$. So long as $M < N/2$, the $O(\sqrt{N})$ behavior holds as an upper bound. And for $M > N/2$, one could either switch to the standard search algorithm, or, if M is not known, artificially double the search space by adding another qubit.

I need to read the Quantum Counting section and polynomial bounds stuff later.

5 Quantum Operations

5.1 Interaction with the environment

If a quantum system is initially uncorrelated with some environment (ie not merely separable, but actually *a product state*), and then some interaction is allowed, after which this environment is discarded (ie implicitly measured but result of measurement unknown), we can describe the dynamics as a quantum operation.

$$\mathcal{E}(\rho) = \mathrm{Tr}_{\mathrm{env}}\left[U\left(\rho \otimes \rho_{\mathrm{env}}\right)U^{\dagger}\right]$$

If we choose a basis $|e_k\rangle$ for the environment space, we can rewrite this trace as

$$\mathcal{E}(\rho) = \sum_k E_k \rho E_k^{\dagger}$$

where

$$E_k = \langle e_k | U | e_0 \rangle$$

This leads to an interpretation of the E_k: each $E_k \rho E_k^{\dagger}$ is the unnormalized state of the system if the environment was left in state $|e_k\rangle$ afterwards. Summing over all the possible environment states (tracing out the environment) gives the expectation state of the system.

We can generalize this to deal with the case of explicit measurement on the environment (ie where the experimenter does have access to the result). If we include a projective measurement $\{P_m\}$, and the measurement result is m, then the final state is

$$\frac{\mathrm{Tr}_E\left(P_m U(\rho \otimes \sigma)U^{\dagger}P_m\right)}{\mathrm{Tr}\left(P_m U\left(\rho \otimes \sigma\right)U^{\dagger}P_m\right)}$$

If we define $\mathcal{E}_m$ equal to the numerator, then the final state is $\mathcal{E}_m\left(\rho\right) / \mathrm{Tr}\left(\mathcal{E}_m\left(\rho\right)\right)$. $\mathrm{Tr}\left(\mathcal{E}_m\left(\rho\right)\right)$ is not 1, but rather the probability of measuring m. If we decompose the environment state $\sigma = \sum_j q_j |j\rangle\langle j|$, this can be put into an operator-sum representation

$$\mathcal{E}_m\left(\rho\right) = \sum_{jk} E_{jk} \rho E_{jk}^{\dagger}$$

where

$$E_{jk} = \sqrt{q_j}\langle e_k | P_m U | j \rangle$$

(Note that the multiple indices on the E_{jk} are due to the generalization to allow the environment to be in a mixed state, not due to the measurement.)

5.2 Axiomatic definition

A quantum operation $\mathcal{E} : Q_1 \to Q_2$ could be defined by three axioms.

1. Since $\mathrm{Tr}\,\mathcal{E}\left(\rho\right)$ is the probability that $\mathcal{E}$ occurs, $0 \leq \mathrm{Tr}\,\mathcal{E}\left(\rho\right) \leq 1$.
2. $\mathcal{E}\left(\sum_i p_i \rho_i\right) = \sum_i p_i \mathcal{E}\left(\rho_i\right)$

3. $\mathcal{E}$ is completely positive. That is, not only is $\mathcal{E}(A)$ positive for any positive A on the system, but $(I \otimes \mathcal{E})(A)$ is positive for any positive A on the tensor product of the system with any other space. The classic example of a positive but not completely positive map is the transpose.

If a map satisfies the above criteria, it can be written in the operator-sum notation, and vice-versa.

There is also a unitary freedom/redundancy in the operator-sum representation worth mentioning. If $\{E_k\}$ form a quantum operation, and U is unitary, than $F_k = \sum_j U_{kj} E_j$ is the same quantum operation. This can be used to show (see write-up for Exercise 8.10) that the operator-sum representation need not contain more than d^2 terms where d is the dimension of the quantum system. (In the case of a mapping from a d-dimension system to a d'-dimension system, the mapping needs no more than dd' terms).

5.3 Trace and Partial Trace

The trace is a quantum operation from a d-dimensional space to a 1-dimensional space:

$$\mathcal{E}(\rho) = \sum_{i=1}^{d} |0\rangle \langle i| \rho |i\rangle \langle 0|$$

The partial trace from QR to Q is also a quantum operation:

$$E_i \left(\sum_j \lambda_j |q_j\rangle |j\rangle \right) = \lambda_i |q_i\rangle$$

Or, bizarrely written:

$$E_i = \langle i_Q|$$

5.4 Geometric Picture

Any quantum operation can be written as an affine map on the Bloch sphere representation:

$$\vec{r} \xrightarrow{\mathcal{E}} M\vec{r} + \vec{c}$$

where M can be decomposed $M = OS$, $O \in SO_3$, S real symmetric. So every quantum operation on a single qubit is a scaling, followed by a proper rotation, followed by a shift.

This picture makes many facts geometricly visualizable. For instance, the purity $\operatorname{Tr} \rho^2$ can be written as $\left(1 + |r|^2\right)/2$. So if an operation always shrinks the Bloch vector, it reduces the purity.

5.5 Specific Channels

5.5.1 Flipping

The bit-flipping channel flips the qubit with probability $1-p$. It can be expressed

$$E_0 = \sqrt{p} I \quad E_1 = \sqrt{1-p} X$$

It compresses the Bloch sphere along the $y-z$ plane, leaving $|+\rangle$ and $|-\rangle$ untouched.

The phase flip channel which probabilistically flips the phase acts analogously (replace X with Y). It shrinks the $x-y$ plane.

The bit-phase flip channel (replace X with Y acts analogously), shrinking the $x-z$ plane. The name makes sense because $Y = iXZ$.

5.5.2 Depolarizing

The depolarizing channel does nothing with probability $1-p$, and depolarizes the qubit–replaces it with $I/2$–with probability p. This can be written in operator-sum form as

$$\mathcal{E}(\rho) = (1-p)\rho + \frac{p}{3}(X\rho X + Y\rho Y + Z\rho Z)$$

5.5.3 Amplitude Damping

Amplitude damping models the dissapation of energy from a quantum system to its environment. One can derive it, for instance, by modelling the interaction of an optical qubit with its (initially empty) environment as a beamsplitter. ($\gamma = \sin^2\theta$ describes the strength of the beamsplitter).

$$E_0 = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{bmatrix} \quad E_1 = \begin{bmatrix} 0 & \sqrt{\gamma} \\ 0 & 0 \end{bmatrix}$$

Notice that the only state left invariant (for non-trivial γ) is $|0\rangle$. This is because we modelled the environment initially empty (zero-temperature). One could model a finite-temperature environment with Generalized Amplitude Damping:

$$E_0 = \sqrt{p}\begin{bmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{bmatrix} \quad E_1 = \sqrt{p}\begin{bmatrix} 0 & \sqrt{\gamma} \\ 0 & 0 \end{bmatrix}$$
$$E_3 = \sqrt{1-p}\begin{bmatrix} \sqrt{1-\gamma} & 0 \\ 0 & 1 \end{bmatrix} \quad E_1 = \sqrt{p}\begin{bmatrix} 0 & 0 \\ \sqrt{\gamma} & 0 \end{bmatrix}$$

I need to go back and write up the quantum process tomography stuff

5.5.4 Phase Damping

The phase damping channel, which provides random phase kicks is equivalent to a phase-flipping channel.

6 Distance Measures for States

6.1 Classical Measures

6.1.1 Trace distance

$$D(p_x, q_x) = \frac{1}{2}\sum_x |p_x - q_x|$$

Trace distance has a nice physical interpretation as the difference in probability of an event S between the distributions q and p, where S is the subset of x's which maximize that difference (in some sense, S is the best event for distinguishing q and p). Explicitly:

$$D(p_x, q_x) = \max_S \left(p(S) - q(S)\right) = \max_S \left(\sum_{x \in S} p_x - \sum_{x \in S} q_x\right)$$

(This can be easily shown by collecting all the x's into S for which $p_x > q_x$)

This also makes it very simple to prove that the trace distance is actually a metric. From it's definition, it's clearly symmetric, and zero iff $p = q$, and,as for the triangle inequality, the above lemma guarantees there is an S such that the following logic holds:

$$D(p_x, q_x) = p(S) - q(S) = (p(S) - w(S)) + (w(S) - q(S)) \leq D(p_x, w_x) + D(w_x, q_x)$$

Note: N&C did not put this quick proof in, but they proved it for the quantum case, and that proof was strikingly analogous to this classical case.

6.1.2 Fidelity

The fidelity is the inner product of $\sqrt{p_x}$ and $\sqrt{q_x}$.

$$F(p_x, q_x) = \sum_x \sqrt{p_x q_x}$$

6.2 Quantum Measures

6.2.1 Trace Distance

$$D(\rho, \sigma) = \frac{1}{2} \operatorname{Tr} |\rho - \sigma|$$

where $|A| = \sqrt{A^\dagger A}$ is the positive square root of $A^\dagger A$.If ρ and σ commute, then they can be simultaneously diagonalized and this will reduce to the classical trace distance.

Conveniently, for qubits, the trace distance between two states is half the Euclidean distance between their Bloch vectors. Also, the trace distance is clearly invariant under unitary transformations.

We can generalize our maximization expression from the classical trace distance to the quantum case:

$$D(\rho, \sigma) = \max_P \operatorname{Tr} \left(P \left(\rho - \sigma\right)\right)$$

The proof of this relation is analogous to that of the classical statement. It uses a trick which is quite handy: $\rho - \sigma$ can be written $Q - S$ where Q and S are positive operators *with orthogonal support*. (This trick is analagous to, in the classical case, carefully grouping the events such that $p_x > q_x$). Using this decomposition, we can write

$$D(\rho, \sigma) = \left(\operatorname{Tr}(Q) + \operatorname{Tr}(S)\right)/2$$

But since ρ and σ are both of unit trace, $\operatorname{Tr}(Q - S) = 0$, ie $\operatorname{Tr}(Q) = \operatorname{Tr}(S)$. Choosing P to be a projector onto the support of Q attains the maximization

and the equality. Just as in the classical case, this above lemma makes it easy to demonstrate that the trace distance is a metric.

A handy physical interpretation is that the quantum trace distance bounds the classical trace distance achievable by measurements:

$$D(\rho, \sigma) = \max_{\{E_m\}} D(p_m, q_m)$$

where that is a maximization over all POVMs $\{E_m\}$. Writing the term inside the maximization as

$$D(p_m, q_m) = \frac{1}{2} \sum_m |\mathrm{Tr}(E_m(\rho - \sigma)|$$

and using the same decomposition trick, one can see that maximization is attained by choosing the E_m to be projectors onto the supports of Q and S.

6.2.2 More properties of trace distance

- Trace-preserving quantum operations are contractive:

$$D(\rho, \sigma) \geq D(\mathcal{E}(\rho), \mathcal{E}(\sigma))$$

- Strong convexity:

$$D(\sum_i p_i \rho_i, \sum_i q_i \sigma_i) \leq \sum_i p_i D(\rho_i, \sigma_i) + D(p_i, q_i)$$

  which also implies joint convexity:

$$D(\sum_i p_i \rho_i, \sum_i p_i \sigma_i) \leq \sum_i p_i D(\rho_i, \sigma_i)$$

  and convexity in either argument:

$$D(\sum_i p_i \rho_i, \sigma) \leq \sum_i p_i D(\rho_i, \sigma)$$

6.2.3 Fidelity

The fidelity is defined as

$$F(\rho, \sigma) = \mathrm{Tr}\, \sqrt{\rho^{1/2} \sigma \rho^{1/2}}$$

In the case where one state is pure,

$$F(\rho, |\psi\rangle \langle\psi|) = \sqrt{\langle\psi| \rho |\psi\rangle}$$

The fidelity is also invariant under unitary transformations.

Uhlmann's theorem gives a beautiful expression for the fidelity. Suppose ρ and σ are two states of Q. If we introduce a second quantum system which is a copy of the first,

$$F(\rho, \sigma) = \max_{|\psi\rangle, |\varphi\rangle} |\langle\psi|\varphi\rangle|$$

where the maximization runs over all purifications of ρ and σ. The proof relies two neat properties of entanglement. First, a purification of ρ can be written

$$|\psi\rangle = (U_R \otimes \sqrt{\rho} U_Q) |m\rangle$$

where $|m\rangle$ is a maximally entangled state. Second, if the indices in the two quantum systems are choosen to match, then we can use

$$\mathrm{Tr}(A^\dagger B) = \langle m| (A \otimes B) |m\rangle$$

where the multiplication on the left can be well defined by matrix multiplication in our choosen bases defined by the indices. Using those properties of entanglement, and one simple property of the trace-absolute value:

$$|\mathrm{Tr}(AU)| \leq \mathrm{Tr}|A|$$

we can easily prove Uhlmann's theorem, and, in fact, if either of the purifications is actually fixed, the statement still works as a maximization over the other purification.

Based on Uhlmann's theorem and on the definition itself, we can easily see that the fidelity ranges from 0 to 1, only acheiving those values when (0) the states have orthogonal support or (1) are equal.

Just as we linked the quantum and classical trace distance via measurement, we do the same for fidelity

$$F(\rho, \sigma) = \min_{\{E_m\}} F(p_m, q_m)$$

where $p_m = \mathrm{Tr}(E_m \rho)$, and $q_m = \mathrm{Tr}(E_m \sigma)$.

6.2.4 Properties of the fidelity and the angle

The fidelity is not a metric; however, it naturally defines one. Viewing the fidelity as the inner product between purifications suggests we could define the angle between states ρ and σ as

$$A(\rho, \sigma) = \arccos F(\rho, \sigma)$$

One can use Uhlmann's theorem to quickly prove this is a metric.

The fidelity acts like an "upside-down" trace-distance, and obeys many analogous properties.

- The fidelity is non-decreasing under trace-preserving quantum operations:

$$F(\mathcal{E}(\rho), \mathcal{E}(\sigma)) \geq F(\rho, \sigma)$$

- The fidelity obeys what we refer to by analogy as strong concavity:

$$F\left(\sum_i p_i \rho_i, \sum_i q_i \sigma_i\right) \geq \sum_i \sqrt{p_i q_i} F(\rho_i, \sigma_i)$$

  which implies joint concavity:

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

and concavity in either argument:

$$F\left(\sum_i p_i \rho_i, \sigma\right) \geq \sum_i p_i F(\rho_i, \sigma)$$

6.2.5 Relationship between distance measures

For many purposes, the fidelity and trace distance are qualitatively equivalent means of characterizing the distance between states. For pure states, they are entirely equivalent; simply geometry shows that

$$D(|\varphi\rangle, |\psi\rangle) = \sqrt{1 - F(|\varphi\rangle, |\psi\rangle)}$$

From this, we can deduce their relationship on mixed states. For any ρ and σ, consider the purifications which satisfy Uhlmann's theorem. Upon those purifications, the fidelity will be the same as on ρ and σ by construction, but the trace distance will be greater or equal. Using that, and the above expression for pure states:

$$D(\rho, \sigma) \leq \sqrt{1 - F(\rho, \sigma)}$$

So a large fidelity implies a small trace distance. The converse can also be shown true, such that we can bound the trace distance:

$$1 - F(\rho, \sigma) \leq D(\rho, \sigma) \leq \sqrt{1 - F(\rho, \sigma)^2}$$

6.2.6 How well do channels preserve quantum information

We can use these distance measures to quantify how well an operation preserves a state. (For convenience, we will work with the fidelity.)

We could say that an operation $\mathcal{E}$ preserves $|\psi\rangle$ if $F(|\psi\rangle, \mathcal{E}(|\psi\rangle\langle\psi|)) \approx 1$, or we could quantify the worst-case behavior of an operation via

$$F_{\min}(\mathcal{E}) = \min_{|\psi\rangle} F(|\psi\rangle, \mathcal{E}(|\psi\rangle\langle\psi|))$$

(That maximization over pure states is equivalent to a maximization over all states because of the concavity of the fidelity). Similarly, we could quantify how well an operation acheives any intended gate with the gate fidelity:

$$F_{\min}(\mathcal{E}) = \min_{|\psi\rangle} F(|\psi\rangle, \mathcal{E}(|\psi\rangle\langle\psi|))$$

Whatever metric, d, that we're using, we can define the error E

$$E(U, \mathcal{E}) = \max_{\rho} d(U\rho U^\dagger, \mathcal{E}(\rho))$$

So long as d satisfies $d(\rho, \sigma) = d(U\rho U^\dagger, U\sigma U^\dagger)$, the triangle inequality will guarantee that errors add no worse than linearly.

6.2.7 Quantum Information

Here we discuss two possible definitions for a quantum information source in the relevant measures of preservation.

First, the *ensemble notion*: a source is a stream of identical systems whose states are independent and identically distributed variables from some fixed set $\{\rho_i\}$ of states with probabilities p_i. The ensemble average fidelity is defined as

$$\bar{F} = \sum_j p_j F(\rho_j, \mathcal{E}(\rho_j))^2$$

where the squared-ness is justified only by convenience at this point.

A second is the *entanglement fidelity*. If the system under discussion is Q, we introduce a fictitious environment R such that the state of RQ is a purification, and define the entanglement fidelity as

$$\begin{aligned} F(\rho, \mathcal{E}) =& F(RQ, R'Q')^2 \\ =& \langle RQ| \left[(I_R \otimes \mathcal{E})(|RQ\rangle \langle RQ|)\right] |RQ\rangle \end{aligned}$$

(Because all purifications of Q are related by a unitary transform on R alone, it doesn't matter which we choose for evaluating that definition.) One attractive property of the entanglement fidelity is that it's easy to calculate, given an operator-sum representation for $\mathcal{E}$:

$$F(\rho, \mathcal{E}) = \sum_i |\mathrm{Tr}\, \rho E_i|^2$$

Of these two notions, the entanglement fidelity is the more stringent measure, as demonstrated by the following two results:

$$F(\rho, \mathcal{E}) \leq [F(\rho, \mathcal{E}(\rho)]^2$$

Intuitively, it is more difficult to preserve a state and its entanglement than just to preserve a state. Secondly, monotonicity of the fidelity under partial trace and convexity of the entanglement fidelity as a function of ρ imply that

$$F(\sum_j p_j \rho_j, \mathcal{E}) \leq \bar{F}$$

So if an operation preserves the entanglement fidelity of a source described by $\rho = \sum_i p_i \rho_i$, then it will automatically do a good job of preserving ensemble average fidelity.

Five easily-provable properties of the entanglement fidelity are promised to come in handy later:

1. $0 \leq F(\rho, \mathcal{E}) \leq 1$. Obvious from definition.
2. F is linear in $\mathcal{E}$. Obvious from definition.
3. For pure states,
$$F(|\psi\rangle, \mathcal{E}) = F(|\psi\rangle, \mathcal{E}(|\psi\rangle \langle\psi|))^2$$
4. $F(\rho, \mathcal{E}) = 1$ iff $\mathcal{E}$ acts as the identity upon the support of ρ.
5. If $\langle\psi| \mathcal{E}(|\psi\rangle \langle\psi|) |\psi\rangle \geq 1 - \eta$ for all $|\psi\rangle$ in the support of ρ, then $F(\rho, \mathcal{E}) \geq 1 - (3\eta/2)$

7 Quantum Error Correction

I already wrote up the section on the three-qubit flip code and phase code and the nine-qubit Shor code for 8.06, so I won't do so again here... for now

7.1 General Theory of QEC

A code C is a subspace. A code is error-correcting against the quantum operation $\mathcal{E}$ if there is a trace-preserving quantum operation $\mathcal{R}$ such that

$$(\mathcal{R} \circ \mathcal{E})(\rho) \propto \rho$$

for any ρ supported by C. The proportionality in place of an equals allows $\mathcal{E}$ to be non-trace-preserving.

If P is the projector onto C, and $\{E_i\}$ represent $\mathcal{E}$, then the existence of such an $\mathcal{R}$ is equivalent to the following error correcting condition:

$$PE_i^\dagger E_j P = \alpha_{ij} P$$

for some Hermitian matrix α. The equivalence proof is quick and intuitive. A certain unitary will diagonalize α; if that unitary is used to recombine the operation elements of $\mathcal{E}$ (remember their unitary freedom) into $\{F_k\}$, then we can write

$$PF_k^\dagger F_l P = d_{kl} P$$

where d is diagonal. A polar decomposition of $F_k P = U_k \sqrt{P F_k^\dagger F_k P} = \sqrt{d_{kk}} U_k P$ shows that F_k just rotates the code into the image $U_k P$, whereas the diagonality of d makes these spaces orthogonal. Then we just apply $U_k^\dagger$ to each image subspace to recover.

$$\mathcal{R}(\sigma) = \sum_k U_k^\dagger P_k \sigma P_k U_k$$

The other direction of the equivalence follows from the observation that the combination of the error and recovery on $P\rho P$ is proportional to $P\rho P$.

It is simple to show from the error correcting condition that if $\mathcal{R}$ and C correct an error process with elements $\{E_i\}$, then they correct any error process with elements formed from other linear combinations of the $\{E_i\}$. Thus we can talk about the space of error operations which are corrected for by a given code.

As an example, if a code corrects against the Pauli matrices on one particular qubit, then it corrects against arbitrary errors on that single qubit. (So it suffices to show that a code corrects against the depolarizing channel.)

7.2 Degeneracy and the Quantum Hamming Bound

One uniquely quantum aspect of error correction is the phenomena of degenerate codes. For instance, in the Shor code, Z_1 and Z_2 both map the code onto the same error space, and reapplication of either will correct the error, whereas with classical codes, no pair of errors on different bits would ever lead to the state.

Intuitively, degeneracy might lead to quantum codes that could store information more compactly than any non-degenerate code; however, no such case has yet been found.

For non-degenerate codes, we can construct a simple bound on how many qubits are necessary for the encoding, just by counting.

Suppose a non-degenerate code encodes k qubits onto n qubits, and can correct arbitrary errors on $t \leq n$ qubits.

Remember that being able to correct for the Pauli matrices on a qubit implies arbitrary error correction for that qubit. If j errors occur, they may occur in $\binom{n}{j}$ possible locations and each may be one of three possible errors. So there is a total of

$$\sum_{j=0}^{t} \binom{n}{j} 3^j$$

possible errors. By non-degeneracy, each error gets its own 2^k-dimensional space of the 2^n-dimensional space of possible states. So by counting,

$$\sum_{j=0}^{t} \binom{n}{j} 3^j 2^k \leq 2^n$$

This is the quantum Hamming bound.

8 Detour: Classical linear codes

A linear code C, which encodes k bits into an n-bit code is represented by an $n \times k$ generator matrix G of ones and zeros with linearly independent columns, which maps messages to their encoded forms $x \to Gx$. While a general code would require $n2^k$ bits to specify, a linear code requires only the nk bits of the matrix.

Complementary to the the generator matrix is the $(n-k) \times n$ parity check matrix H with linearly independent rows, whose k-dimensional kernel is the code. By Gaussian elimination, any parity check matrix can be put into standard form $[A|I_{n-k}]$, with A an $(n-k) \times k$ matrix. The corresponding generator matrix is annihilates H from the right: $G = \left[\frac{I_k}{-A}\right]$.

The parity check matrix makes syndrome measurement trivial. If an error takes $y \to y' = y + e$, then $Hy' = He$, and H acts invertibly on the space of errors because it has linearly independent rows.

Let the Hamming distance $d(x, y)$ between two words be the number of bits at which those words differ, and let the Hamming weight of x is $\mathrm{wt}(x) = d(x, 0)$. It follows that $d(x, y) = \mathrm{wt}(x+y)$. If the probability of a bit flip is less than $1/2$, then the most likely correction to a corrupted word is the nearest in Hamming distance. Thus a code with a minimum Hamming distance of $2t+1$ can correct errors on up to t bits.

One more construction is promised to come in handy. If C is a $[n, k]$ code (with generator G, parity check H), then the dual of C, written $C^\perp$ is the code with generator $H^\perp$ and parity check $C^\perp$. It's simple to show that the dual code is simply the subspace orthogonal to C. Weakly and strongly self-dual are $C \subseteq C^\perp$ and $C = C^\perp$, respectively.

9 Calderbank-Shor-Steane codes

We can actually now form a quantum code from two classical codes. Suppose C_1 and C_2 are $[n, k_1]$ and $[n, k_2]$ classical codes, with both C_1 and $C_2^{\perp}$ correcting t errors and $C_2 \subset C_1$.

We now define the quantum code $CSS(C_1, C_2)$. It is the space spanned by the states $|x + C_2\rangle$, where

$$|x + C_2\rangle = \frac{1}{\sqrt{|C_2|}} \sum_{y \in C_2} |x + y\rangle$$

It's easy to see that each coset of C_2 in C_1 produces one unique vector of the code, so the dimension is the number of cosets $|C_1| / |C_2| = 2^{k_1 - k_2}$ so $CSS(C_1, C_2)$ is an $[n, k_1 - k_2]$ code. N&C steps through the error correction procedure for this code.