5.3 Quantum data compression

Schumacher invented the quantum analogue of classical source coding. The classical source coding theorem states that a source with entropy H can be compressed to H bits per source letter. Schumacher's quantum source coding theorem states that a quantum source with von Neumann entropy S can be compressed to S qubits per source output.

In quantum source coding, a transmitting party, Alice, would like to encode an n-length message coming out of a quantum source into a k-qubit quantum codeword, with k/n as small as possible, so that a receiver, Bob, when provided with the k-qubit quantum codeword can recover the n-length message with as high a fidelity as desired. Notice that, unlike classical source coding, we do not require an exact reproduction of the original source state, because that is in general impossible. So, a notion of fidelity is necessarily introduced in quantum source coding. There are various possible formulas for fidelity. We will use the following one. The fidelity of reconstructing a state $|u\rangle$ as state $|v\rangle$ is defined as

$$F = |\langle v|u\rangle|^2.$$

Since quantum opearations are typically stochastic, this definition is too narrow. In general, we will allow that Alice's message $|u\rangle$ be reconstructed by Bob as some state $|v_x\rangle$ with some probability q_x. In this case, the fidelity is defined as

$$F = \sum_x q_x |\langle v_x|u\rangle|^2 = \langle u|\rho'|u\rangle,$$

where

$$\rho' = \sum_x q_x |v_x\rangle\langle v_x|.$$

5.3.1 An example

Let the source generate the states

$$|\psi_1\rangle = |0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad |\psi_2\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

with probabilities 1/2 each.

The case $n = 1$

Let's first consider source coding for length $n = 1$ messages. If Alice can send $k = 1$ qubit to Bob, then she can send the source output direcly with fidelity $F = 1$.

It's instructive to consider the case $k = 0$ also, in which Alice does not send anything to Bob.

What is the best that Bob can then do maximize the fidelity? A little thought shows that an optimal strategy for Bob is to generate a guess $|\phi\rangle$ such that

$$F = p_1|\langle\phi|\psi_1\rangle|^2 + p_2|\langle\phi|\psi_2\rangle|^2 = \langle\phi|\rho|\phi\rangle$$

is maximized, where

$$\rho = \frac{1}{2}|\psi_1\rangle\langle\psi_1| + \frac{1}{2}|\psi_2\rangle\langle\psi_2| = \begin{pmatrix} 3/4 & 1/4 \\ 1/4 & 1/4 \end{pmatrix}$$

is the density matrix for the source. It turns out the solution of the compression problem lies in the spectral decomposition of ρ,

$$\rho = \sum_x \lambda_x |u_x\rangle\langle u_x|,$$

where λ_x are the eigenvalues and $|u_x\rangle$ are the eigenvectors of ρ. Now, it is easily seen that

$$F = \langle\phi|\rho|\phi\rangle = \sum_x \lambda_x |\langle\phi|\psi_x\rangle|^2 \leq \lambda_{max},$$

where λ_{max} is the largest eigenvalue of ρ, with equality iff $|\phi\rangle$ is chosen as an eigenvector with eigenvalue λ_{max}. For the example here, the eigenvectors and eigenvalues are

$$|u_1\rangle = \begin{pmatrix} \cos(\pi/8) \\ \sin(\pi/8) \end{pmatrix}, \quad \lambda_1 = \cos^2(\pi/8)$$

$$|u_2\rangle = \begin{pmatrix} \sin(\pi/8) \\ -\cos(\pi/8) \end{pmatrix}, \quad \lambda_2 = \sin^2(\pi/8).$$

The optimal guess is therefore $|\phi\rangle = |u_1\rangle$ and the fidelity is $F = \cos^2(\pi/8) = .8535$.

The case $n = 3$, $k = 2$

Now, Alice will encode 3 source qubits into 2 qubits. What is her optimal strategy? One possibility is that she sends the first 2 qubits directly and lets Bob guess the third qubit. This results in fidelity $F = 1 \times 1 \times .8535 = .8535$. This method can be improved as follows.

The spectral decomposition of the density matrix ρ^3 for three source qubits is given by

$$\rho^3 = \sum_{x_1, x_2, x_3} \lambda_{x_1}\lambda_{x_2}\lambda_{x_3} |u_{x_1}, u_{x_2}, u_{x_3}\rangle\langle u_{x_1}, u_{x_2}, u_{x_3}|$$

where $x_i = 1, 2$, and $\{|u_i\rangle, \lambda_i\}$ are the eigenvectors and eigenvalues of ρ. Consider now the 4-dimensional subspace Λ spanned by the eigenvectors with the largest eigenvalues, namely, $|u_1, u_1, u_1\rangle$, $|u_1, u_1, u_2\rangle$, $|u_1, u_2, u_1\rangle$, and $|u_2, u_1, u_1\rangle$. The idea is to encode the source output into this high-probability subspace. Let E denote the projector onto this subspace. To encode a source output triplet $|\psi\rangle = |\psi_1, \psi_2, \psi_3\rangle$, Alice performs a projective measurement with operators E and $I - E$. With probability $\langle\psi|E|\psi\rangle$, the output of the measurement equals

$$|\psi'\rangle = \frac{E|\psi\rangle}{\sqrt{\langle\psi|E|\psi\rangle}},$$

and with the residual probability, it equals

$$|\psi''\rangle = \frac{(I - E)|\psi\rangle}{\sqrt{\langle\psi|(I - E)|\psi\rangle}}.$$

This projection can be conveniently carried out by means of a unitary transformation U that carries the four eigenvectors that span Λ into the standard basis vectors $|000\rangle$, $|010\rangle$, $|100\rangle$, and $|110\rangle$. Alice begins the compression algorithm by first applying the transformation

$$|\psi\rangle \to U|\psi\rangle = U|\psi'\rangle + U|\psi''\rangle.$$

Notice that the third qubit of the state $U|\psi'\rangle$ is always $|0\rangle$ and the third qubit of $U|\psi''\rangle$ is always $|1\rangle$. Now, Alice measures the third qubit only. If the result is 0, Alice sends the first two (unmeasured) qubits to Bob and discards the third qubit. If the result is 1, Alice sends the two junk qubits, to be specific, say, the first 2 qubits of $U|u_1, u_1, u_1\rangle$. This turns out to be a convenient choice that simplifies fidelity calculations, but otherwise has no significance.[1]

[1]The choice of what is sent as junk does not affect asymptotic performance as the length parameter n

Upon receipt of the two qubits, Bob appends a qubit $|0\rangle$ as the third qubit and performs the operation U^{-1} on the three qubits. At this point, it can be verified with a little effort that, Bob has obtained a state with density

$$\rho' = E|\psi\rangle\langle\psi|E + (1 - \langle\psi|E|\psi\rangle)\rho_{junk}$$

where $\rho_{junk} = |u_1, u_1, u_1\rangle\langle u_1, u_1, u_1|$. The resulting fidelity is

$$F \quad = \quad |\langle\psi|E|\psi\rangle|^2 + (1 - \langle\psi|E|\psi\rangle)|\langle\psi|u_1, u_1, u_1\rangle|^2$$

We now calculate

$$\begin{aligned}
\langle\psi|E|\psi\rangle &= |\langle u_1, u_1, u_1|\psi\rangle|^2 + |\langle u_1, u_1, u_2|\psi\rangle|^2 + |\langle u_1, u_2, u_1|\psi\rangle|^2 + |\langle u_2, u_1, u_1|\psi\rangle|^2 \\
&= \cos^6(\pi/8)3\cos^4(\pi/8)\sin^2(\pi/8) \\
&= .6219 + 3 \times .1067 = .9419.
\end{aligned}$$

Thus, the overall fidelity is obtained as

$$F = (.9417)^2 + (1 - .9417)(.6219) = .9234,$$

which improves upon the simpler method discussed earlier. In this example, the fidelity turned out to be independent of the source state $|\psi\rangle$; in general, the fidelity depends on the source state, and we consider the fidelity averaged over all states.

5.3.2 The general case

A quantum source is a source that generates qubits from an ensemble $\{|\psi_x\rangle,\ p_x\}$ where the states are not necessarily orthogonal. The density of a single qubit by the source is

$$\rho = \sum_x p_x|\psi_x\rangle\langle\psi_x|$$

The source is memoryless, so n consecutive qubits by the source, called a "message of length n," has density

$$\rho^n = \rho \otimes \cdots \otimes \rho.$$

tends to infinity. It does however affect the performance for small examples such as this one. Here we are more interested in illustrating the general idea than finding the exact optimal code.

Let $\mathcal{H}_n$ denote the Hilbert space of length-n messages. This space has dimension N^n, where $1 \leq N \leq 2$ is the number of non-zero eigenvalues of ρ.

The essense of quantum data compression problem is to find high-probability subspaces with as low a dimensionality as possible. More precisely, we seek a subsapce Λ_n of $\mathcal{H}_n$ such that, if E_n denotes the projection operator onto Λ_n,

$$P(\Lambda_n) = \mathrm{tr}(\rho^n E) \to 1,$$

and, at the same time we wish to minimize the normalized dimensionality

$$\frac{1}{n}\log_2(\mathrm{dim}\Lambda_n).$$

Schumacher solved this problem and showed that the smallest possible value of the normalized dimensionality is given by the von Neumann entropy $S(\rho)$.

The result gains an operational meaning in the context of quantum data compression. Given a message $|\psi_i\rangle$ of length n, we may carry out a projective measurement with projectors $\{E_n, 1-E_n\}$. With probability $\langle\psi_i|E|\psi_i\rangle$, the measurement result is

$$|\psi_i'\rangle = \frac{E|\psi_i\rangle}{\sqrt{\langle\psi_i|E|\psi_i\rangle}}$$

and $\psi_i'\rangle$ can be stored or sent over a quantum channel using roughly $nS(\rho)$ qubits. With remaining probability, the measurement result is in the orthogonal complement of Λ_n, and in this case, it matters little how we choose to store or communicate in place of $|\psi_i\rangle$. So, the encoding operation can be summarized by the transformation

$$|\psi_i\rangle\langle\psi_i| \to \rho_i' = E|\psi_i\rangle\langle\psi_i|E + \rho_{i,junk}\langle\psi_i|(1-E)|\psi_i\rangle$$

The fidelity of such representation, averaged over all messages, is given by

$$\begin{aligned}
F &= \sum_i p_i\langle\psi_i|\rho_i'|\psi_i\rangle \\
&= \sum_i p_i\langle\psi_i|E|\psi_i\rangle\langle\psi_i|E|\psi_i\rangle + \sum_i p_i\langle\psi_i|\rho_{i,junk}|\psi_i\rangle\langle\psi_i|(1-E)|\psi_i\rangle \\
&\geq \sum_i p_i|\langle\psi_i|E|\psi_i\rangle|^2 \\
&= \sum_i p_i|\langle\psi_i|E^2|\psi_i\rangle|^2
\end{aligned}$$

$$
\begin{aligned}
&= \sum_i p_i ||E|\psi_i\rangle||^4 \\
&\geq \sum_i p_i (2\langle\psi_i|E|\psi_i\rangle - 1) \\
&= 2\mathrm{tr}(\rho^n E) - 1 \\
&\geq 2(1-\epsilon) - 1 \\
&= 1 - 2\epsilon,
\end{aligned}
$$

where the second inequality from above is based on the fact that $x^2 \geq 2x - 1$ for any real number x.

The question that remains is how to find such a subspace Λ_n. For this, one needs to consider the spectral decomposition of the density matrix

$$\rho = \lambda_1 |u_1\rangle\langle u_1| + \lambda_2 |u_2\rangle\langle u_2|$$

where λ_i are the eigenvalues and $|u_i\rangle$ are the eigenvectors of ρ. The eigenvectors of the product density ρ^n are given by tensor products of the eigenvectors of ρ. These can be conveniently enumerated by a vector $x = (x_1, \ldots, x_n)$, $x_i = 1, 2$, so that

$$|u_x\rangle = |u_{x_1}\rangle \otimes \cdots \otimes |u_{x_n}\rangle$$

is an eigenvector of ρ^n with eigenvalue

$$\lambda_x = \lambda_{x_1} \cdots \lambda_{x_n}.$$

We now define Λ_n as the subspace spanned by the eigenvectors of ρ^n with eigenvalues λ_x satisfying

$$2^{-n(S-\delta)} \geq \lambda_x \geq 2^{-n(S+\delta)}$$

where $\delta > 0$ is a parameter and $S = S(\rho)$. We note that

$$S(\rho) = H(\lambda_1, \lambda_2) = -\lambda_1 \log \lambda_1 - \lambda_2 \log \lambda_2.$$

We now invoke the asymptotic equipartition property (AEP), discussed earlier. Each eigenvector of ρ^n may be thought of as a random variable with probability equal to its eigenvalue. The set Λ_n defined above then corresponds to the typical set of eigenvectors with respect to parameter δ. The AEP states that, given any $\epsilon > 0, \delta > 0$, for n sufficiently large, the typical set has probability

$$P(\Lambda_n) = \mathrm{tr}(\rho^n E) > 1 - \epsilon$$

and dimension

$$2^{n(S+\delta)} \geq \dim(\Lambda_n) \geq (1-\epsilon)2^{n(S-\delta)}.$$

This completes the argument showing that representation of ρ^n using a subspace Λ_n of dimension $\leq n(S+\delta)$ and with fidelity $\geq 1 - 2\epsilon$ is possible for any given $\delta > 0$, $\epsilon > 0$, provided n is sufficiently large.

For actual storage or communication of an element of Λ_n using exactly $d = \dim(\Lambda_n)$ qubits, we may apply a unitary transformation that maps each eigenvector $|u_x\rangle$ in the subspace Λ_n to a vector of the form

$$U|u_x\rangle = |u_x'\rangle|0_{rest}\rangle$$

where $|u_x'\rangle$ has dimension d and $|0_{rest}\rangle$ is an all-zero state of dimension $n-d$. This transforms each element of the subspace Λ_n to a vector in which the trailing $n-d$ qubits are in state $|0_{rest}\rangle$. The compressed version is obtained by discarding the qubits in state $|0_{rest}\rangle$. In order to decode the compressed copy, we first append $|0_{rest}\rangle$ and apply the inverse operation U_{-1}.

The converse argument

We now wish to determine if the compression factor attained by the above choice of Λ_n is the best possible value. Consider now an arbitrary subspace Λ_n of dimension d. Let the compression algorithm map a state ψ_i to a density ρ_i' such that

$$\rho_i' = \sum_i \alpha_i |a_i\rangle\langle a_i|$$

where $|a_i\rangle$ are basis vectors of Λ_n. The fidelity of this representation is given by

$$\begin{aligned}
F_i &= \langle\psi_i|\rho_i'|\psi_i\rangle \\
&= \sum_i \lambda_i \langle\psi_i|a_i\rangle\langle a_i|\psi_i\rangle \\
&\leq \sum_i \langle\psi_i|a_i\rangle\langle a_i|\psi_i\rangle \\
&\leq \langle\psi_i|E_n|\psi_i\rangle
\end{aligned}$$

The average fidelity is given by

$$F = \sum_i p_i F_i \leq \sum_i p_i \langle \psi_i | E_n | \psi_i \rangle = \mathrm{tr}(\rho^n E_n).$$

The trace term is upperbounded by the sum of the largest d eigenvalues of ρ^n, which goes to zero in n unless $d > nS(\rho)$. This completes the converse.