Seminar on Quantum Compression

Ofer Shayevitz, April 2006

I Introduction - Classical Compression

Data compression is a branch of *Information Theory*, a field studying the mathematical theory of communications, founded by Claude E. Shannon in a landmark paper [1] dated 1948.

Definition I.1. *An information source in general is a random process. A discrete memoryless source (DMS) is an i.i.d process over a finite discrete alphabet $\mathcal{X}$, with letter probabilities $p(x)$, i.e.,*

$$P_r(x_1,\ldots,x_n) = P_r(x^n) = \prod_{i=1}^{n} p(x_i)\,, \quad x_i \in \mathcal{X}$$

We will discuss the DMS case only. The fundamental question of classical compression is: **How many bits per source letter are required on the average to reliably represent the source?** We shall define what we mean by *reliable* in a moment.

Definition I.2. *A block code $\mathcal{C}$ for a source X is a quadruplet $\mathcal{C} = (E, D, R, F)$ where*[1]

$$\begin{aligned} \textit{Encoder}: &\quad E : \mathcal{X}^n \mapsto \mathcal{Y}^k \\ \textit{Decoder}: &\quad D : \mathcal{Y}^k \mapsto \mathcal{X}^n \\ \textit{Compression Rate}: &\quad R = \frac{k \log |\mathcal{Y}|}{n} \ \ [\textit{bit/letter}] \\ \textit{Fidelity}: &\quad F(\mathcal{C}) = P_r\Big\{x^n \in \mathcal{X}^n \ : \ D(E(x^n)) = x^n\Big\} \end{aligned}$$

where $\mathcal{Y}$ is a finite code alphabet, and the probability is taken w.r.t. the source's distribution.

Definition I.3. *A rate R is called* **attainable** *for a source X if for any $\varepsilon > 0$ there exists a rate R code $\mathcal{C}$ for the source with fidelity $F(\mathcal{C}) > 1-\varepsilon$. The optimal compression rate for the source is the infimum of all achievable rates.*

Remark I.1. Block codes are a special case and in general a code may be of variable length. However, in terms of achievable compression rates considering block codes is enough, and more general coding schemes can only improve upon the trade-off between delay and fidelity.

[1] All logarithms are taken to the base of 2.

Example I.1. Let X be a binary DMS with alphabet $\mathcal{X} = \{0, 1\}$ and probability distribution $p(1) = p\,, p(0) = 1 - p$. The probability of a block x^n of n source letters is

$$P_r(x^n) = p^{n_1}(1-p)^{n-n_1}$$

where n_1 is the number of 1's in x^n. For a large n we expect that $n_1 \sim np$. This can be quantified using, say, Chebyshev inequality as follows:

$$\begin{aligned}
&\mathbb{E}(n_1) = np\,, \quad Var(n_1) = np(1-p) \\
&P_r\Big\{|n_1 - np| > \delta n\Big\} \leq \frac{Var(n_1)}{(\delta n)^2} = \frac{p(1-p)}{\delta^2 n} \quad \underset{n\to\infty}{\longrightarrow} \; 0
\end{aligned}$$

Define a *Typical Set* A_n^δ as follows:

$$A_n^\delta = \Big\{x^n \in \mathcal{X}^n \;:\; \Big|\sum x_i - np\Big| < \delta n\Big\}\,, \quad P_r\big\{A_n^\delta\big\} \underset{n\to\infty}{\longrightarrow} 1\,, \forall \delta > 0$$

Now, the probability of any sequence $x^n \in A_n^\delta$ is given by

$$P_r(x^n) = p^{\,n(p+\alpha)}(1-p)^{n(1-p-\alpha)}$$

for some $|\alpha| < \delta$. Taking the logarithm we get:

$$\log P_r(x^n) = n(p+\alpha)\log p + n(1-p-\alpha)\log(1-p)$$

Define the *binary entropy function*

$$h(p) = -\Big(p\log(p) + (1-p)\log(1-p)\Big)$$

we have that

$$\log P_r(x^n) = -n\left(h(p) + \alpha \log\frac{1-p}{p}\right)$$

and so

$$2^{-n\left(h(p)+\delta\left|\log\frac{1-p}{p}\right|\right)} \leq P_r(x^n) \leq 2^{-n\left(h(p)-\delta\left|\log\frac{1-p}{p}\right|\right)}$$

or simply

$$2^{-n\left(h(p)+\widehat{\delta}\right)} \leq P_r(x^n) \leq 2^{-n\left(h(p)-\widehat{\delta}\right)}$$

where $\widehat{\delta}$ was implicitly defined. Now let $\varepsilon, \delta > 0$. Then for n large enough

$$\begin{aligned}
&1 - \varepsilon \leq P_r\big\{A_n^\delta\big\} \leq 1 \\
&\big|A_n^\delta\big| \leq \frac{1}{2^{-n\left(h(p)+\widehat{\delta}\right)}} = 2^{n\left(h(p)+\widehat{\delta}\right)} \\
&\big|A_n^\delta\big| \geq \frac{1-\varepsilon}{2^{-n\left(h(p)-\widehat{\delta}\right)}} = (1-\varepsilon)2^{n\left(h(p)-\widehat{\delta}\right)}
\end{aligned}$$

Loosely speaking, there is a relatively small set of $\sim 2^{nh(p)}$ *typical sequences* emitted each attained with probability $\sim 2^{-nh(p)}$, with total probability arbitrarily close to 1.

Now the way to compression is short! First, arrange the typical sequences in A_n^δ in some order (e.g. lexicographic), i.e., construct a bijective mapping

$$I : A_n^\delta \mapsto \{0,1\}^{\lceil \log |A_n^\delta| \rceil}$$

and consider the following block code:

$$\begin{aligned}
E(x^n) &= \begin{cases} I(x^n) & x^n \in A_n^\delta \\ 00\ldots0 & o.w. \end{cases} \\
D(y) &= I^{-1}(y) \\
R &= \frac{\lceil \log |A_n^\delta| \rceil}{n} \leq \frac{n(h(p)+\widehat{\delta})+1}{n} = h(p) + (\widehat{\delta} + \frac{1}{n}) \\
F &= P_r\{A_n^\delta\} \geq 1-\varepsilon
\end{aligned}$$

Thus any rate above the binary entropy $h(p)$ is achievable! In fact, no rate below $h(p)$ is achievable as we shall momentarily see, which makes $h(p)$ the optimal compression rate for this source.

Remark I.2. Note that for $0 \leq h(p) \leq 1$ and attains the lower bound for $p = 0, 1 \rightarrow$ no bits are required to reliably describe a deterministic source. The upper bound is attained for $p = \frac{1}{2} \rightarrow$ a uniform source cannot be compressed.

Remark I.3. F=1 can be simply attained by using variable length coding - whenever outside the typical set transmit the entire source block as a codeword, and add a flag bit at the beginning of each codeword to indicates whether inside or outside the typical set.

This property of the existence of a (small) typical set with constant probability sequences that has probability close to 1, is called the *Asymptotic Equipartition Property* (AEP). The AEP also holds for general DMS (and is true in fact for any stationary ergodic source), where the size of the typical set is related to the *entropy* of the source.

Definition I.4. *Let X be a DMS over an alphabet $\mathcal{X}$ with letter probability $p(x)$. The entropy of the source is defined as*

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x)$$

Notice that for the special case of a binary source with probability p we have that $H(X) = h(p)$.

Theorem I.1 (AEP). *Let X be a DMS with entropy $H = H(X)$, and let $\varepsilon, \delta > 0$. For any n large enough there exists a set $A_n^{\varepsilon,\delta} \in \mathcal{X}^n$ satisfying*

$$\begin{aligned} 1-\varepsilon \leq & P_r\left(A_n^{\varepsilon,\delta}\right) \leq 1 \\ (1-\varepsilon)2^{n(H-\delta)} \leq & \left|A_n^{\varepsilon,\delta}\right| \leq 2^{n(H+\delta)} \\ 2^{-n(H+\delta)} \leq & P_r(x^n) \leq 2^{-n(H-\delta)}, \quad \forall x^n \in A_n^{\varepsilon,\delta} \end{aligned}$$

Furthermore, let $B_n \in \mathcal{X}^n$ be a sequence of sets with size $|B_n| < 2^{n(H-\delta)}$. Then for any n large enough $P_r(B_n) < \varepsilon$.

Proof. First part similar to the binary example. Second part is easily derived by looking at the intersection of B_n with typical sets. □

Theorem I.2 (Classical Source Coding). *Let $\varepsilon, \delta > 0$. For any n large enough there exists a block code $\mathcal{C}$ with rate $R < H(X) + \delta$ and fidelity $F > 1 - \varepsilon$. Conversely, for all n large enough, any block code $\mathcal{C}$ with rate $R < H(X) - \delta$ has fidelity $F < \varepsilon$.*

Proof. **Achievability:** Similar to the binary example. **Converse:** Any block code with rate $R < H - \delta$ can describe no more than $2^{n(H-\delta)}$ source sequences. According to the AEP, this size of a sequence set has a vanishing probabilistic volume, which results in a vanishing fidelity.

□

II Quantum Compression

First described in [2] and with a simpler proof in [3]. The proofs consider only a *Unitary Decoder* for the converse. A converse for the most general setting was given in [4].

Definition II.1. *A (discrete memoryless) quantum source is a probability distribution over a finite ensemble of pure states $\{|a_i\rangle, p_i\}_{i=1}^M$, with a density matrix $\rho = \sum_{i=1}^M p_i |a_i\rangle\langle a_i|$.*

The fundamental question of quantum compression is: **How many qubits per source state are required on the average to reliably represent the source?** We shall define what we mean by *reliable* in a moment.

Consider a quantum source $A = \{|a_i\rangle, p_i\}_{i=1}^M$ over a d-dimensional Hilbert space $\mathcal{H}_d$, represented by its density matrix ρ. A general *encoder* for the source is a mapping from input states to density matrices:

$$|a_i\rangle \to \Omega_i$$

where the density matrices Ω_i are over a k-dimensional Hilbert space $\mathcal{H}_k$ (the dimension is reduced = compression). A general *decoder* is a mapping

$$\Omega_i \to \omega_i$$

where ω_i are density matrices over $\mathcal{H}_d$ again.

How reliable is the reconstruction? The input $|a_i\rangle$ in reconstituted as the mixed state ω_i whose probability to pass a yes/no test as being $|a_i\rangle$ is given by $\langle a_i|\omega_i|a_i\rangle$.

Definition II.2. *The fidelity of the coding scheme is defined as*

$$F = \sum_{i=1}^{M} p_i \langle a_i|\omega_i|a_i\rangle$$

For example, if the coding scheme does nothing then $\omega_i = |a_i\rangle\langle a_i|$ and then $F = 1$. In general we have $0 \leq F \ \leq 1$.

Similarly to the classical setting, we can consider *blocks* of source states by aggregating n source's inputs together. The equivalent source A_n has M^n states in the Hilbert space $\mathcal{H}_{d^n}$, and is represented by the density matrix $\rho^{\otimes n}$. The probabilities of the different states are merely multiplications of probabilities of states in A, and are denoted $p_i^{(n)}$. Coding and Fidelity are defined in a similar manner for A_n. In coding A_n we use $\log k$ qubits per block, and so *compression rate* of the scheme is given by

$$R = \frac{\log k}{n} \ [qubits/state]$$

Example II.1. Consider a binary source A with states $\{|0\rangle, |+\rangle\}$ with probabilities $\{p, 1-p\}$. The source A_2 has eight states

$$\{|00\rangle, |0+\rangle, |+0\rangle, |++\rangle\}$$

with probabilities

$$\{p_i^{(2)}\} = \{p^2, p(1-p), (1-p)p, (1-p)^2\}$$

Example II.2. Consider a binary source A with states $\{|0\rangle, |+\rangle\}$ with uniform probabilities $\{\frac{1}{2}, \frac{1}{2}\}$. Thinking classically, it seems there is no way to compress the source even when looking at blocks, since all the blocks have the same probability. However, there is a significant difference from the classical setting. For instance, when guessing the output of a classical binary uniform source, the best fidelity possible is $F = \frac{1}{2}$. However, for the quantum source in question, our best guess would obviously be the vector $|0\rangle + |+\rangle$ (with normalization), which will provide a much higher fidelity of $F = \cos^2 \frac{\pi}{8} \approx 0.8536$.

Let us then look at the source A in the basis $|0'\rangle = |0\rangle + |+\rangle$ and $|1'\rangle = |0\rangle - |+\rangle$ (with normalization). We can think of $|0'\rangle$ as spanning a "likely" 1-dimensional subspace for the

source's states, and of $|1'\rangle$ as spanning an "unlikely" 1-dimensional subspace. Extending that notion, we can look at A_n in the orthogonal basis $\mathcal{B} = \{|0'\rangle, |1'\rangle\}^{\otimes n}$. Express

$$|0\rangle = \cos\frac{\pi}{8}|0'\rangle + \sin\frac{\pi}{8}|1'\rangle$$
$$|+\rangle = \cos\frac{\pi}{8}|0'\rangle - \sin\frac{\pi}{8}|1'\rangle$$

Then for any source state $|\psi\rangle = |\psi_1\rangle|\psi_2\rangle \dots |\psi_n\rangle$ (where each qubit is in either $|0\rangle$ or $|+\rangle$) we have that

$$\begin{aligned}|\psi\rangle &= \left(\cos\frac{\pi}{8}|0'\rangle \pm \sin\frac{\pi}{8}|1'\rangle\right) \dots \left(\cos\frac{\pi}{8}|0'\rangle \pm \sin\frac{\pi}{8}|1'\rangle\right) \\ &= \sum_{\phi\in\{0',1'\}^n} \pm \left(\sin\frac{\pi}{8}\right)^{n_1(\phi)} \left(\cos\frac{\pi}{8}\right)^{n-n_1(\phi)} |\phi\rangle\end{aligned}$$

where $n_1(\phi)$ is the number of $1'$ in ϕ. Therefore, the projection of **any source state** $|\psi\rangle$ on a basis vector $|\phi\rangle \in \mathcal{B}$ is

$$|\langle\psi|\phi\rangle|^2 = \left(\sin^2\frac{\pi}{8}\right)^{n_1(\phi)} \left(\cos^2\frac{\pi}{8}\right)^{n-n_1(\phi)} = \lambda^{n_1(\phi)}(1-\lambda)^{n-n_1(\phi)}$$

where we set $\lambda = \sin^2\frac{\pi}{8}$ and $1-\lambda = \cos^2\frac{\pi}{8}$. But this is precisely the probability of a classical binary sequence with $n_1(\phi)$ ones, produced by a DMS with probabilities $(1-\lambda, \lambda)$!

Loosely speaking, as we have seen in the classical setting, there exists $\sim 2^{nh(\lambda)}$ typical sequences with total probability close to 1. This means that there exists $\sim 2^{nh(\lambda)}$ vectors in $\mathcal{B}$ that span a *typical subspace* with the property that the source states are projected into it with probability close to 1. This suggests a simple encoding scheme that projects A_n into the typical subspace, thus reliably describing the source using $\sim nh(\lambda)$ qubits per block, resulting in a compression rate of $R \sim h(\lambda)$.

The derivations in the example above are implicitly related to the density matrix ρ of the source. The states $|0'\rangle, |1'\rangle$ are just the eigenstates of ρ, and its eigenvalues are just the same $(1-\lambda, \lambda)$ as above, i.e.,

$$\rho = (1-\lambda)|0'\rangle\langle 0'| + \lambda|1'\rangle\langle 1'|$$

This clarifies the following definition.

Definition II.3. *The Von-Neumann entropy of a quantum source with density matrix ρ is defined as*

$$S(\rho) \stackrel{\triangle}{=} -\sum \lambda_i \log\lambda_i = -tr\ \rho\log\rho$$

where λ_i are the eigenvalues of ρ.

The Von-Neumann entropy is generally smaller than the Shannon entropy (of the generating probability distribution), and equal to it only for orthogonal states.

Theorem II.1 (Quantum Source Coding). *Let $\varepsilon, \delta > 0$. For all n sufficiently large, there exists a block coding scheme attaining a compression rate $R < S(\rho) + \delta$ and a fidelity $F > 1 - \varepsilon$. Conversely, For all n sufficiently large, any block coding scheme with a compression rate $R < S(\rho) - \delta$ attains a fidelity $F < \varepsilon$.*

Proof. **Achievability:** The eigenvalues of the density matrix satisfy $\lambda_i \geq 0$ and $\sum \lambda_i = 1$, and thus constitute a probability distribution. Therefore, in terms of a density matrix we can think of the source A as an essentially classical source A', producing the eigenstates $|\lambda_i\rangle$ of ρ with probability λ_i. The Shannon entropy of A‘ is equal to the Von-Neumann entropy of A, i.e., $H(A') = S(\rho)$.

Similarly, the source A_n can be though of as an essentially classical source A'_n, producing orthonormal eigenstates of $\rho^{\otimes n}$ that are tensor products the eigenstates $|\lambda_i\rangle$, with probabilities that are multiplications of the corresponding λ_i's. For any $\varepsilon, \delta > 0$, we have (by the AEP) that for n large enough there are no more than $2^{n(S(\rho)+\delta)}$ *typical eigenstates* of $\rho^{\otimes n}$ whose total probability (sum of eigenvalues) exceeds $1 - \frac{\varepsilon}{2}$. We denote the subspace spanned by these eigenstates by $\Lambda = \Lambda(n, \varepsilon, \delta)$, and call it the *Typical Subspace*. Following that, the quantum compression scheme works as follows:

Encoder: (works on A_n)

1. Makes a measurement that projects the input state into either Λ or $\Lambda^{\perp}$.
2. If the measurement indicates that the state was projected into $\Lambda^{\perp}$, substitutes the state with a predefined state $|0\rangle \in \Lambda$.
3. Applies a unitary operator U that does the following:

$$|\phi\rangle \in \Lambda \xrightarrow{U} |\psi\rangle|0_{rem}\rangle\,,$$

   where $|\psi\rangle$ is a state with $n(S(\rho) + \delta)$ qubits, and $|0_{rem}\rangle = |0\rangle \otimes \ldots \otimes |0\rangle$ are the remaining qubits.
4. The first $n(S(\rho) + \delta)$ are sent or stored.

Decoder:

1. Adds $|0_{rem}\rangle$ ancilla qubits.
2. Applies $U^{\dagger}$ to $|\psi\rangle|0_{rem}\rangle$ to retrieve $|\phi\rangle \in \Lambda$.

The rate of the proposed scheme is $R = S(\rho) + \delta$ as required, so let us now evaluate its fidelity. Let $|a_i\rangle$ denote an input state of A_n, and express it as

$$|a_i\rangle = \alpha_i|\ell_i\rangle + \beta_i|m_i\rangle\,, \quad |\ell_i\rangle \in \Lambda\,,\ |m_i\rangle \in \Lambda^\perp$$

Then the density matrix associated with the decoder's output is

$$\omega_i = |\alpha_i|^2|\ell_i\rangle\langle\ell_i| + |\beta_i|^2|0\rangle\langle 0|$$

The per state fidelity can be bounded then by

$$\langle a_i|\omega_i|a_i\rangle = |\alpha_i|^2|\langle a_i|\ell_i\rangle|^2 + |\beta_i|^2|\langle a_i|0\rangle|^2 \geq |\alpha_i|^2|\langle a_i|\ell_i\rangle|^2 = |\alpha_i|^4 \geq 2|\alpha_i|^2 - 1$$

Defining Λ also as the projector onto the typical subspace, the scheme fidelity is bounded correspondingly

$$F = \sum p_i^{(n)}\langle a_i|\omega_i|a_i\rangle \geq 2\sum p_i^{(n)}|\alpha_i|^2 - 1 = 2\sum p_i^{(n)}\langle a_i|\Lambda|a_i\rangle - 1 = 2\sum tr\left(p_i^{(n)}|a_i\rangle\langle a_i|\Lambda\right) - 1$$
$$= 2\,tr\rho^{\otimes n}\Lambda - 1 \geq 2(1-\frac{\varepsilon}{2}) - 1 = 1 - \varepsilon$$

as desired.

The complexity of the scheme lies mainly in the unitary operation U. This is essentially a classical computation which orders typical sequences, and can be performed using classical coding techniques such as *enumerative coding* [5], performed in a reversible quantum-mechanical manner. Coding techniques using $O(n^3)$ elementary 2-3 quantum gates were first introduced in [6], and an improved algorithm with quasi-linear complexity $O(n(\log^4 n)\log\log n)$ was described in [7].

Converse: We shall only prove the converse for a unitary decoder, meaning that we assume the decoder adds ancilla and performs a unitary operation. This means that the density matrices ω_i are all embedded in a common subspace Γ of dimension $k = 2^{n(S(\rho)-\delta)}$. Therefore, ω_1 has an orthonormal basis $|\xi_1\rangle, \ldots, |\xi_k\rangle$ in Γ. Defining $\Pi_i = |a_i\rangle\langle a_i|$ we can write

$$\omega_1 = \sum_{j=1}^{k} q_j|\xi_j\rangle\langle\xi_j|\,, \quad q_j \geq 0$$

and so the per state fidelity is

$$\langle a_1|\omega_1|a_1\rangle = \sum_{j=1}^{k} q_j\langle a_1|\xi_j\rangle\langle\xi_j|a_1\rangle \leq \sum_{j=1}^{k}|\langle a_1|\xi_j\rangle|^2 = tr\Pi_1\Gamma$$

This is true for any ω_i and thus the average fidelity is bounded by

$$F \leq \sum p_i^{(n)}\,tr\Pi_i\Gamma = tr\rho^{\otimes n}\Gamma$$

Now, introducing the basis $\{|e_i\rangle\}_{i=1}^{d^n}$ of the eigenstates of $\rho^{\otimes n}$ with eigenvalues $\{\mu_i\}_{i=1}^{d^n}$, we can rewrite the above as

$$F \leq \sum_{i=1}^{d^n} \mu_i \, tr|e_i\rangle\langle e_i|\Gamma = \sum \mu_i \langle e_i|\Gamma|e_i\rangle$$

But $0 \leq \langle e_i|\Gamma|e_i\rangle \leq 1$ and $\sum\langle e_i|\Gamma|e_i\rangle = tr\Gamma = k$, and therefore the fidelity F is bounded above by the sum of the kth largest eigenvalues of $\rho^{\otimes n}$. However, since $k = 2^{n(S(\rho)-\delta)}$ we know from the AEP that for n large enough this sum is smaller than ε, proving the converse. □

Remark II.1. Note that only the density matrix related to the source is important in terms of compression, and not the **actual states** emitted by the source. This means for instance that a coding scheme for one source performs the same for any other source with the same density matrix. This justifies the definition of a quantum source simply as a density matrix.

Remark II.2. It can be shown that the converse holds for arbitrary decoders [4], so there is no gain using a non-unitary decoder in terms of compression rates. However, improvement in fidelity for a given coding scheme may be obtained by more general decoders [4].

Remark II.3. Notice that in the proof of the converse we have made no assumptions regarding the encoding process. This means that even if the encoder knows the actual quantum state being coded, this knowledge cannot be used to decrease the achievable compression rate. The setting where the encoder is aware of the input quantum states is called *visible coding* while the setting where it has no such knowledge is called *blind coding*.

The following is a generalized definition of a quantum source.

Definition II.4. *A (discrete memoryless) mixed quantum source is a probability distribution over a finite ensemble of density matrices* $\{\rho_i, p_i\}_{i=1}^{M}$.

The fidelity of a coding scheme for a mixed quantum source is measured in terms of the average Bures-Uhlmann fidelity between density matrices. In this case it is easily seen that a rate higher than $S(\sum p_i\rho_i)$ is generally attainable. For instance, consider the case where $p_1 = 1$ and $p_i = 0$ for all $i > 1$. In this case, the encoder does not need to send any qubits to the decoder and a perfect fidelity is attainable, thus a rate $R = 0$ is attainable, which is different from $S(\rho_1)$. The best compression possible for this setting in general is nevertheless still an open problem. It is known that the Holevo bound $S(\sum p_i\rho_i) - \sum p_i S(\rho_i)$ is a lower bound [8], and also that in this case there is a possibility of a compression gain for the visible setting w.r.t. the blind setting [9].

III Universal Classical Compression

Consider the case where the probability distribution of the DMS X is unknown, but the source is known to have entropy $H(X) < H_U$. As it turns out, a compression rate of H can still be guaranteed for this entire *family of sources*.

Definition III.1. *For a binary alphabet $\mathcal{X} = \{0, 1\}$, define the* **type** T_p^n *to be the set of all binary strings of length n whose empirical distribution is $(1-p, p)$, i.e.,*

$$T_p^n = \left\{ x^n \in \mathcal{X}^n \,:\, \sum x_i = np \right\}$$

This definition is easily extended to non-binary alphabets, where p is replaced by a probability distribution.

It is easily seen that there are precisely $n+1$ different types. For a non-binary alphabet, the number of types can be bounded from above by $(n+1)^{|\mathcal{X}|-1}$. The important point is that **the number of types is polynomial in n.**

The size of a type is easily bounded from above by the following simple argument. for every $x^n \in \mathcal{X}^n$ we have that $P_r(x^n) = 2^{-nh(p)}$ and therefore

$$1 \geq P_r(T_p^n) = \sum_{x^n \in \mathcal{X}^n} P_r(x^n) = \sum_{x^n \in \mathcal{X}^n} 2^{-nh(p)} = |T_p^n| 2^{-nh(p)}$$

and therefore

$$|T_p^n| \leq 2^{nh(p)}$$

It can also be shown that $|T_p^n| \geq (n+1)^{-|\mathcal{X}|} 2^{nh(p)}$ by similar arguments or by using the Stirling approximation.

Example III.1. Continuing the classical binary source example, we assume now that all we know is that $H(X) < H_U$. Let $\varepsilon, \delta > 0$, and let

$$p_0 = \max\{p \,:\, h(p) \leq H_U + \delta \,, p \leq \frac{1}{2}\}$$

Now define the sets

$$B_0^\delta(H_U) = \bigcup_{k=0}^{\lceil np_0 \rceil} T^n_{\frac{k}{n}} \,, \quad B_1^\delta(H_U) = \bigcup_{k=n-\lfloor np_0 \rfloor}^{n} T^n_{\frac{k}{n}} \,, \quad B^\delta(H_U) = B_0^\delta(H_U) \cup B_1^\delta(H_U)$$

Obviously, the set $B^\delta(H_U)$ contains all the typical sets A_n^δ of any binary source with $H(X) < H_U$. Therefore, a block code the encodes only the sequences in $B^\delta(H_U)$ will faithfully reconstruct any source in that family. The size of $B^\delta(H_U)$ is bounded using the fact that there are no more than $n+1$ types, and the type with maximal size corresponds to p_0:

$$|B^\delta(H_U)| \leq (n+1) \left| T^n_{\frac{\lceil np_0 \rceil}{n}} \right| \leq (n+1) 2^{n(H_U + \delta)} = 2^{n(H_U + \delta + \frac{\log(n+1)}{n})}$$

and so the rate of H_U bits per source letter can be approached arbitrarily close. In the non-binary case the same method holds since there is only a polynomial number of types.

Remark III.1. In the classical setting, it is actually possible to do much better, and approach the entropy of each source in the family. For the family of binary sources, this can be done by counting the number of ones in a length n sequence, sending this number using $\log n$ bits, and then sending the index of the sequence inside its type (arranged according to, say, a lexicographical order). Since the size of the type is $\leq 2^{nh(p_{emp})}$ where p_{emp} is the empirical fraction of ones in the sequence, and as the $\log n$ bits used to identify the type are negligible w.r.t. n, we approach a compression rate of $h(p_{emp})$ bits per letter. Since $p_{emp} \to p$ in the limit of large blocks, the entropy of the source is approached. Much more efficient algorithms exists [10] which work "on the fly" and approach the compression limit. These algorithms also apply to sources with memory and even to "individual" sources with no underlying probabilistic mechanism (for such sources the known Lempel-Ziv compression scheme [10] was shown to asymptotically "beat" the best finite-state compression scheme of any order).

IV Universal Quantum Compression

The extension of the universal scheme for compression of classical sources with bounded entropy is not straightforward [11]. Consider a family of quantum sources with a density matrix ρ and a Von-Neumann entropy $S(\rho) < S_U$. Can we guarantee a compression rate of S_U qubits per state for any source in this family?

Instead of projecting into a **single typical subspace**, one should use a subspace containing the **union of all possible typical subspaces** for all the sources with entropy $S(\rho) < S_U$, to guarantee fidelity approaching one for all the sources in the family. What is the dimension of the minimal subspace containing the union of all those typical subspaces?

Example IV.1. Let us first consider a family of binary quantum sources A with a common density matrix ρ and eigenvalues $\lambda, 1-\lambda$, satisfying $S(\rho) < S_U$. This problem is equivalent to the classical setting by looking at the classical family of sources A' emitting eigenstates of ρ with probability $(\lambda, 1-\lambda)$ and possessing a Shannon entropy $h(\lambda) < S_U$. These family of sources can be compressed to S_U bits per letter by the method of types as in the previous subsection. In the quantum setting this means we project on a subspace of eigenstates of $\rho^{\otimes n}$ that correspond to **types** with a bounded entropy, and since the number of types is polynomial in n we can attain a compression rate of S_U qubits per state.

The above procedure can be used for a broader family of quantum sources sharing the same eigenstate structure, i.e., sources whose density matrices commute. But this is still far from what we aim for.

Let us discuss the general binary quantum setting, where the only restriction on the sources is $S(\rho) < S_U$. This discussion is easily generalized to sources over larger alphabets.

Let B_0 be some orthonormal basis of $\mathcal{H}_2$. Any other orthonormal basis B can be obtained from B_0 by applying some unitary transformation U to the basis vectors. Now, think of such a B as the eigenbasis of some density matrix ρ, and let $T^n(B)$ be the set of $2^{n(S_u+\delta)}$ eigenstates of $\rho^{\otimes n}$ which span the subspace used for compressing all the sources that share this eigenbasis and have entropy $S(\rho) < S_U$ (just as depicted in the previous example). We immediately see that the elements in $T(B)$ can be obtained from those in $T(B_0)$ by a unitary transformation. This is informally stated as

$$T(B) = T(UB_0) = U^{\otimes n}T(B_0)$$

for some unitary matrix U. Now define the subspace

$$\Gamma = span\Big\{U^{\otimes n}\boldsymbol{\phi} \,:\, U \; unitary\,, \boldsymbol{\phi} \in T(B_0)\Big\}$$

Compression by projecting on the subspace Γ will reliably represent any source with entropy $S(\rho) < S_U$. Therefore, it is the *dimension* of Γ that we now seek. To that end we shall use the following relaxation:

$$\Gamma \subseteq span\Big\{A^{\otimes n}\boldsymbol{\phi} \,:\, A \in \mathbb{C}^{2\times 2}\,, \boldsymbol{\phi} \in T(B_0)\Big\}$$

For any fixed $\boldsymbol{\phi}$ define

$$\Gamma_\phi = span\Big\{A^{\otimes n}\boldsymbol{\phi} \,:\, A \in \mathbb{C}^{2\times 2}\Big\}$$

and so

$$\dim\Gamma \leq \sum_{\phi\in T(B_0)} \dim\Gamma_\phi$$

We shall now prove that

$$\dim\Gamma_\phi \leq (n+1)^4$$

and since $|T(B_0)| \leq 2^{n(h(S_U)+\delta)}$ we shall conclude that

$$\dim\Gamma \leq (n+1)^4\; 2^{n(h(S_U)+\delta)}$$

which means a compression rate of S_U qubits per state is asymptotically achievable for the entire family of sources, as required.

To prove that above, we introduce a new definition and prove a Lemma.

Definition IV.1. *Let $\mathcal{L}$ be some linear vector space. The symmetric subspace of $\mathcal{L}^{\otimes n}$ is the space $SYM(\mathcal{L})$ of vectors which are invariant under any permutation of the positions in the tensor product.*

Example IV.2. The symmetric subspace of $\mathcal{H}_2^{\otimes 3}$ is spanned by the four vectors

$$|000\rangle, |001\rangle + |010\rangle + |100\rangle, |011\rangle + |101\rangle + |110\rangle, |111\rangle$$

Notice that these vectors correspond to binary types.

Lemma IV.1. *Let $d = \dim \mathcal{L}$. Then*

$$\dim SYM(\mathcal{L}) = \binom{n+d-1}{d}$$

and $SYM(\mathcal{L})$ is also spanned by all the vectors of the form $\psi^{\otimes n}$.

Proof. For simplicity we shall prove the above only for $d = 2$, the proof is rather easily extended. As we have seen in the example preceding the lemma, a basis for the symmetric subspace corresponds to vectors which are sums over *types*. These vectors are all orthogonal and so the dimension of the symmetric subspace is simply the number of types, which in the binary case is just $n + 1$ as required.

Now, from the definition of the symmetric subspace it is straightforward that $\psi^{\otimes n} \in SYM(\mathcal{L})$ for any $\psi \in \mathcal{L}$. We now show that every element in the symmetric subspace can be expressed as a linear combination of such vectors. Denote by $\{|t_0\rangle, \ldots, |t_n\rangle\}$ the "type vectors" that span $SYM(\mathcal{L})$ as in the example, i.e., $|t_k\rangle$ is the sum of all vectors with k ones. Any $|\phi\rangle \in SYM(\mathcal{L})$ can be expressed as

$$|\phi\rangle = \sum_{j=0}^{n} a_j |t_j\rangle$$

Now for some set of scalars $\{\alpha_i\}_{i=0}^n$, consider vectors of the form

$$|\psi_i\rangle^{\otimes n} = (|0\rangle + \alpha_i|1\rangle)^{\otimes n} = \sum_{\phi \in \{0,1\}^n} \alpha_i^{n_1(\phi)} |\phi\rangle = \sum_{k=0}^{n} \alpha_i^k |t_k\rangle$$

Let us try to represent $|\phi\rangle$ using the vectors $|\psi_i\rangle^{\otimes n}$:

$$|\phi\rangle = \sum_{j=0}^{n} a_j |t_j\rangle = \sum_{i=0}^{n} b_i |\psi_i\rangle^{\otimes n} = \sum_{i=0}^{n} b_i \sum_{k=0}^{n} \alpha_i^k |t_k\rangle = \sum_{k=0}^{n} |t_k\rangle \sum_{i=0}^{n} b_i \alpha_i^k$$

which has the following matrix representation

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ \alpha_0 & \alpha_1 & \cdots & \alpha_n \\ \alpha_0^2 & \alpha_1^2 & \cdots & \alpha_n^2 \\ \vdots & \vdots & & \vdots \\ \alpha_0^n & \alpha_1^n & \cdots & \alpha_n^n \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_n \end{bmatrix} = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix}$$

But the matrix is just the Van-Der-Monde matrix, which is always invertible if none of the α_i's are equal, so the representation is feasible which completes our proof. □

Following the lemma, we can write

$$\Gamma_\phi = span\Big\{ A^{\otimes n}\boldsymbol{\phi} \,:\, A \in \mathbb{C}^{2\times 2} \Big\} = span\Big\{ B\boldsymbol{\phi} \,:\, B \in SYM(\mathbb{C}^{2\times 2}) \Big\}$$

Define a linear map L from $SYM(\mathbb{C}^{2\times 2})$ to Γ_ϕ by

$$L(B) = B\boldsymbol{\phi}\,, \quad \forall B \in SYM(\mathbb{C}^{2\times 2})$$

This map is onto Γ_ϕ, and since a linear map cannot increase dimension we have that

$$\dim \Gamma_\phi \leq \dim SYM(\mathbb{C}^{2\times 2})$$

But since $\dim(\mathbb{C}^{2\times 2}) = d = 4$ we have from the lemma that

$$\dim SYM(\mathbb{C}^{2\times 2}) = \binom{n+d-1}{d-1} \leq (n+1)^{d^2} = (n+1)^4$$

and thus

$$\dim \Gamma_\phi \leq \dim SYM(\mathbb{C}^{2\times 2}) \leq (n+1)^4$$

which yields

$$\dim \Gamma \leq (n+1)^4 \; 2^{n(h(S_U)+\delta)}$$

as required.

There are better universal quantum compression techniques that imitate the classical ones and compress the source to its entropy, without prior knowledge of the latter [12].

Bibliography

[1] C.E. Shannon, "A Mathematical theory of communication," *Bell Sys. Tech Journal,* 27: 379-432, 623-656, 1948.

[2] B. Schumacher, "On Quantum Coding," *Phys. Rev. A*, 1993.

[3] R. Jozsa and B. Schumacher, "A new proof of the quantum noiseless coding theorem," *J. Modern Opt.,* vol. 41, pp. 2343-2349, 1994.

[4] H. Barnum, C. A. Fuchs, R. Jozsa and B. Schumacher, "General Fidelity Limit for Quantum Channels," *Phys. Rev. A,* 54, 4707-4711, 1996.

[5] T.M. Cover, "Enumerative Source Coding," *IEEE Trans. Inform. Theory*, vol. IT-19, no. 1, pp. 73-77, Jan. 1973.

[6] R. Cleve and D.P. DiVincenzo, "Schumacher's quantum data compression as a quantum computation," *Physical Review A,* Vol. 54, No. 4, pp. 2636-2650, 1996.

[7] J.H. Reif and S.Chakraborty, "Efficient and Exact Quantum Compression," accepted to the *Journal of Information and Computation*, 2006.

[8] H. Barnum, C.M. Caves, C.A. Fuchs, R. Jozsa and B.W. Schumacherm, "On quantum coding for ensembles of mixed states," *Technical Report CSTR-00-015,* Department of Comp. Science, University of Bristol, August 2000

[9] M. Horodecki, "Limits for compression of quantum information carried by ensembles of mixed states," *Physical Review A,* Vo. 57, 1998.

[10] J. Ziv and A. Lempel, "Compression of individual sequences via variable-rate coding," *IEEE Trans. Info. Theory,* vol. IT-24, pp. 530-536, Septembet 1978.

[11] R. Jozsa, M. Horodecki, P. Horodecki and R. Horodecki, "Universal Quantum Information Compression," *Physical Review Letters,* vol. 81, 1998.

[12] R. Jozsa and S. Presnell, "Universal quantum information compression and degrees of prior knowledge," *Proc. R. Soc. Lond. A,* vol. 459, pp. 3061-3077, October 2003.