Topics in Quantum Information — **U. Waterloo CO 781, Spring 2004**
Ashwin Nayak (Instructor) — **Niel de Beaudrap (Scribe)**
Information Theory (continued). Error correction. — **June 3, 2004**

1 Mutual information

We would like to define some concept of the information shared between two parties, based on how the probability distributions of the two parties' data relate to one another. For this, we introduce the concept of the mutual information between of random variables.

1.1 Classical Mutual Information

In classical information theory, we would like the following properties to hold for such a measure, for any two random variasbles X and Y :

- If X is independent of Y, the mutual information should be 0.
- If X and Y are equal, then the mutual information should be maximized (and in particular, would be the information content of either X or Y considered alone).

To that end, we introduce the following definition.

Definition 1 *The mutual information $I(X:Y)$ of two random variables X and Y is given by*

$$I(X:Y) \;=\; H(X) + H(Y) - H(XY) \tag{1}$$

We can verify easily that $I(X:Y)$ has the following properties:

(i). If X and Y are independent, $I(X:Y) = 0$, because $H(XY) = H(X) + H(Y)$. Similarly, if X and Y are always equal, $I(X:Y) = H(X)$, becase $H(XY) = H(X) = H(Y)$. (Thus, $I(X:Y)$ has the property we originally wanted.)

(ii). If X is a uniformly random bit, and Y has the distribution $\{(X,p),\ (1-X\ ,\ 1-p)\}$ for some $p > \frac{1}{2}$, then

$$I(X:Y) \;=\; 1 - H(p). \tag{2}$$

(iii). Generalizing the last point, if $\Pr(Y = X) = p$, then

$$I(X:Y) \;\geq\; H(X) - (1-p)\log(|\mathcal{X}| - 1) - H(p) \tag{3}$$

where $\mathcal{X}$ is the sample space for the variable X. (This result is called *Fano's Inequality*.)

(iv). A corollary to the previous point: If X is uniformly distributed over $\{0,1\}^n$, and $\Pr(X = Y) = p$, then

$$I(X:Y) \;\geq\; n - (1-p)n - 1 \;\geq\; pn - 1 \tag{4}$$

which we can obtain by applying the inequality $n \geq \log(2^n - 1)$ to Fano's inequality.

(v). For any two random variables X and Y, we know that $I(X:Y) \leq H(X)$ holds, because $H(XY) - H(Y) \geq 0$; and similarly for Y.

(vi). Suppose that we have a *Markov chain* of random variables, $X \to Z \to Y$: that is, Z is calculated from X alone and Y is calculated from Z alone. Then $I(X:Y) \leq I(X:Z)$. (This is called the *data processing inequality.*)

(vii). A corollary to the previous several points: if Z (as in the preceding proprty) is taken from $\{0,1\}^m$, then $I(X:Y) \leq I(X:Z) \leq H(Z) \leq m$. Combining this with (iv), we have $m \geq pn - 1$. (Compare this with the bound from an earlier lecture, which from $p \leq 2^{m-n}$ deduces that $m \geq n + \log(p)$.)

1.2 Quantum Mutual Information

We can define quantum mutual information in an analogous way:

Definition 2 *If XY is a (system having a) mixed state over a Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$, then the* quantum mutual information $I_q(X:Y)$ *is defined by*

$$I_q(X:Y) \;=\; S(X) + S(Y) - S(XY) \tag{5}$$

where $X = \mathrm{Tr}_{\mathcal{H}_B}(XY)$ and $Y = \mathrm{Tr}_{\mathcal{H}_A}(XY)$.

Properties of the quantum mutual information:

(i). $0 \leq I_q(X:Y)$, as follows from the subadditivity of von Neumann entropy. However, unlike classical mutual information, the quantum information shared between two systems may be larger than the entropy in either part, and best general upper bound we have is $I_q(X:Y) \leq 2S(X)$, and $I_q(X:Y) \leq 2S(Y)$. This follows form the Araki-Lieb inequality.

(ii). If XZ is a (system having a) mixed state over $\mathcal{H}_A \otimes \mathcal{H}_B$, and Y is the result of some general quantum operation on Z (i.e., the result of a sequence of operations

$$XZ \xmapsto{\substack{\text{add some}\\ \text{ancilla}}} XZ \otimes A \xmapsto{I_X \otimes U_{ZA}} XZY' \xmapsto{\substack{\text{trace out}\\ \text{some qubits}}} XZY$$

which do not involve the subsystem X), then $I_q(X:Y) \leq I_q(X:ZY') = I_q(X:Z)$. (This is analogous to the data processing inequality of classical information, and is usually called the *monotonicity of mutual information* in the quantum setting.)

The last item is a more general form of a famous result:

Theorem 3 (Holevo) *Let X be a classical random variable (i.e. a mixture over standard basis states), Z be the result of an encoding $x \mapsto \rho_x$, and Y be the outcome of a measurement of Z along with some ancilla. Then*

$$I(X:Y) \;\leq\; I_q(X:Z) \;=\; S\left(\sum_x p(x)\rho_x\right) - \sum_x p(x) S(\rho_x)\;. \tag{6}$$

Remark: *The quantity on the right hand side of Equation 6,*

$$\chi\left(\Big\{(p(x)\,,\,\rho_x)\Big\}_x\right) \;=\; S\left(\sum_x p(x)\rho_x\right) - \sum_x p(x) S(\rho_x)\;, \tag{7}$$

is called the Holevo chi quantity.

Equation 6 follows easily from the observation that

$$\rho_{XZ} = \begin{bmatrix} p(x_1)\rho_{x_1} & & & \\ & p(x_2)\rho_{x_2} & & \\ & & \cdots & \\ & & & p(x_n)\rho_{x_n} \end{bmatrix} \tag{8}$$

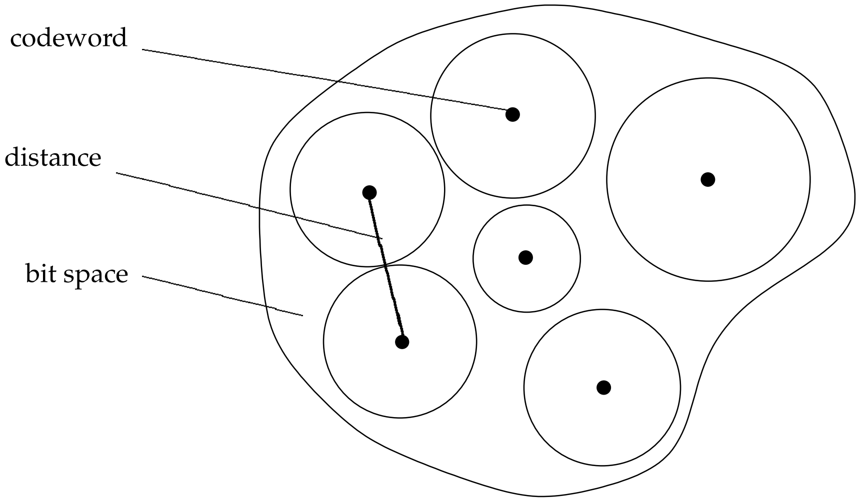
(with off diagonal values all set to zero), and the fact that $S(XZ) = S(X) + \sum_x p(x)S(\rho_x)$.

Corollary 4 *If X is n bits long and Z is m qubits long, then $I_q(X:Y) \leq S\left(\sum_x p(x)\rho_x\right) \leq m$.*

Along with the last property of classical mutual information given above, we see that the Holevo bound implies a bound of $m \geq pn - 1$ on the number of qubits required to transmit n classical bits with probability at least p. As we saw in the beginning of the course, the Nayak bound gives the tighter relationship $m \geq n - \log_2 \frac{1}{p}$.

2 Calderbank-Shor-Steane (CSS) codes

First, we breifly revisit classical error correcting codes. In a classical code, there are codewords which represent valid datastrings, which are separated by a significant distance. A data string which is close to a codeword is presumed to be the result of a small error: by dividing the space into regions surrounding the codewords and correcting strings by projecting them to the nearest codeword, we can correct small errors.



One particularly nice way of performing error correction is with a linear code.

Definition 5 *A (binary) linear $[n, k, d]$ code C is a linear subspace of dimension k of $\mathbb{Z}_2{}^n$, such that the Hamming weight of any non-zero codeword is greater or equal to d.*

We can use special linear codes to define CSS codes: we will see which properties are necessary in the following lecture. What will motivate that special property is an interesting effect that Hadamard operations have on superpositions of codewords from linear codes. Suppose that we have a linear code C: we then define

Definition 6 *The* dual space *of a (binary) linear code C is the subspace $C^\perp \leq \mathbb{Z}_2{}^n$ such that*

$$C^\perp \;=\; \left\{ w \,\middle|\, [\forall u \in C] : w \cdot u = 0 \right\} \tag{9}$$

Interesting properties of the dual space:

(i). $\{0\}^\perp \;=\; \mathbb{Z}_2{}^n$

(ii). $\dim C^\perp \;=\; n - k \;=\; n - \dim C$

(iii). $\left(C^\perp\right)^\perp \;=\; C$

Then, if we define the state

$$|C\rangle \;=\; \frac{1}{\sqrt{2^k}} \sum_{u \in C} |u\rangle \;, \tag{10}$$

then we have

$$H^{\otimes n} |C\rangle \;=\; \frac{1}{\sqrt{2^{n-k}}} \sum_{w \in C^\perp} |w\rangle \;=\; |C^\perp\rangle \;. \tag{11}$$

Proof : We proceed directly:

$$\begin{aligned} H^{\otimes n} |C\rangle \;&=\; \frac{1}{\sqrt{2^k}} \sum_{u \in C} \sum_{x} \frac{1}{\sqrt{2^n}} (-1)^{x \cdot u} |x\rangle \\ &=\; \frac{1}{\sqrt{2^{n+k}}} \sum_{x} \left[\sum_{u \in C} (-1)^{x \cdot u} \right] |x\rangle \end{aligned}$$

Call the amplitude of x α_x : this is the value $\left\langle x \middle| \left(H^{\otimes n}\right) \middle| C \right\rangle$. Now, for $x \in C^\perp$, we know that $x \cdot u \;=\; 0$ for all $u \in C$, so we see that $\alpha_x \;=\; \frac{1}{\sqrt{2^{n-k}}}$. But then, we have

$$\sum_{x \in C^\perp} |\alpha_x|^2 \;=\; 1 \;;$$

then, none of the components for $x \notin C^\perp$ can have any amplitude, so $\alpha_x \;=\; 0$ for $x \notin C^\perp$. ■