CO781 / QIC 890:

Theory of Quantum Communication

Topic 3, part 3

Joint Typicality

Classical communication through noisy classical channel
Shannon's noisy channel coding theorem

Proving the capacity expression for classical channels

References:

Cover & Thomas, Chapter 8

Recall from last lecture:

Def: [achievable rate]

For a channel N, a rate R is achievable if $\exists$ sequence of $(\lfloor 2^{nR} \rfloor, n)$ codes C_n s.t. $P_e(C_n) \to 0$ as $n \to \infty$.

Def: The capacity of N, $C(N)$, is the supremum over achievable rates.

Thm: Shannon's noisy coding theorem

$$C(N) = \max_{p(x)} I(X;Y)$$

where $p(x,y) = \underbrace{p(x)}_{\text{optimized over}} \cdot \underbrace{p(y|x)}_{\text{Given by the channel}}$

Proving the capacity theorem

(1) Prove a <u>direct coding theorem</u>:

In this case, given any $p(x)$, show $\exists$ codes achieving the rate $I(X:Y)$ such that $P_e(C_n) \to 0$.

This gives $C(N) \geq \max_{p(x)} I(X:Y)$.

(2) Prove a <u>converse</u>:

For any achievable R, $R \leq \max_{p(x)} I(X:Y)$,

which gives $C(N) \leq \max_{p(x)} I(X:Y)$.

Here, the upper bound (converse) on the capacity matches the lower bound achieved by codes -- so we know the capacity expression.

Direct coding theorem: fix any $p(x)$.

* Need to show $\exists (M, n)$ codes C_n s.t
  - rate $\frac{1}{n} \log M \geq I(X;Y) - \delta_n$, $\delta_n \to 0$
  - error $P_e(C_n) \to 0$

* Shannon: no need to find these codes.

Instead, $\forall n$, generate C_n by a random process

Show: $\mathbb{E}_{C_n} \bar{P}_e(C_n) \to 0$

(1) Main step of the proof
Randomized argument

error for one code (Cn) averaged over all code words

average over the code Cn

Then: $\exists \tilde{C}_n$ s.t. $\bar{P}_e(\tilde{C}_n) \to 0$

(2) This is immediate from (1)
fix one such code $\tilde{C}_n$

Then: $\exists \tilde{\tilde{C}}_n$ s.t. $P_e(\tilde{\tilde{C}}_n) \to 0$

(3) From $\tilde{C}_n$, "expunge" the bad codewords to reduce error ..

Will see detail, and why rate unchanged!

Step (1):

(averaged over codes, and codewords in each code)

To bound $\mathbb{E}_{C_n} \overline{P_e}(C_n)$:

① Given any n, M, $p(x)$, we generate C_n as follows:

For $i = 1, \ldots, M$

$j = 1, \ldots, n$

draw x_{ij} i.i.d $\sim p(x)$. ← where $p(x)$ appears

particular code that has been chosen,

The C_n, consists of the M codewords:

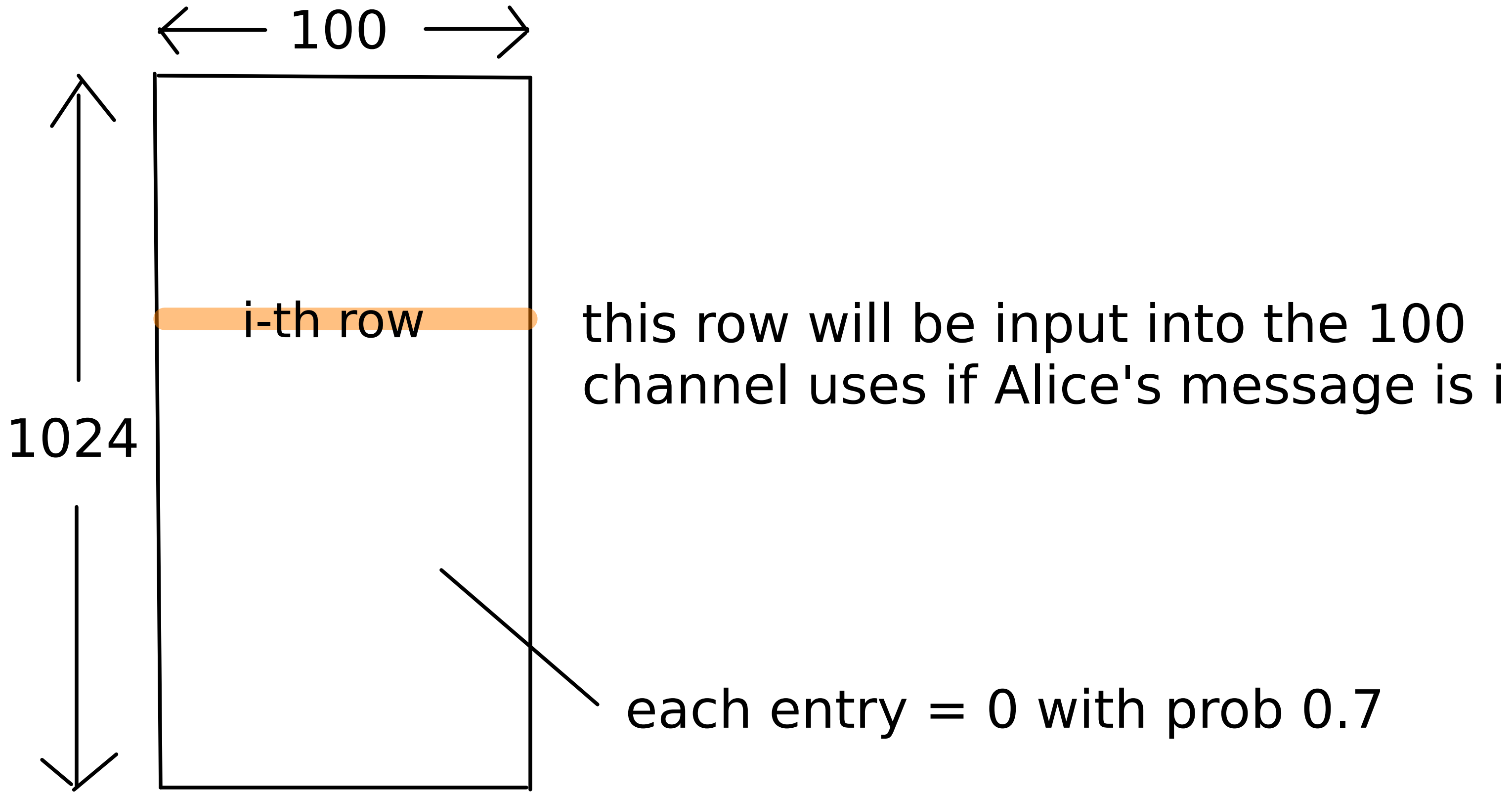
$c_1 = x_{11}\, x_{12} \cdots x_{1n}$

$c_2 = x_{21}\, x_{22} \cdots x_{2n}$

$\vdots$

$c_M = x_{M1}\, x_{M2} \cdots x_{Mn}$

e.g., binary input, p(0) = 0.7, p(1) = 0.3, n = 100, M = 1024.



(2) The code C_n (i.e. $c_1, c_2, \ldots, c_M$) is told to Alice & Bob

(So they know which code has been chosen.)

(3) A message i is drawn randomly from $\{1, 2, \ldots, M\}$ by Alice

(Actually, most of the argument works for any arbitrary i.)

(4) Alice sends c_i through $N^{\otimes n}$. (i.e. $\mathcal{E}_n(i) = c_i$.)

the i-th row in the n-by-M matrix

(5) Bob receives output $Y^n \sim$

$$\Pr(y^n | c_i) = \Pr(y_1 | x_{i1}) P(y_2 | x_{i2}) \cdots P(y_n | x_{in})$$

(6) Bob uses the decoding map $\mathcal{D}_n$:

If $\exists!\, j$ s.t. $c_j y^n \in A_{n,\delta}$, output j

Else "ERR"

"Joint typicality decoding" is suboptimal compared to maximum likelihood decoding, but asymptotically still capacity achieving and easier to analyse.

In the above procedure, what is the probability of error?

Averaged over the choice of code Cn:

c_i is n iid draws $\sim p(x)$

$c_i y^n$ is n iid draws $\sim p(x,y) = p(x)\cdot p(y|x)$

for n large enough

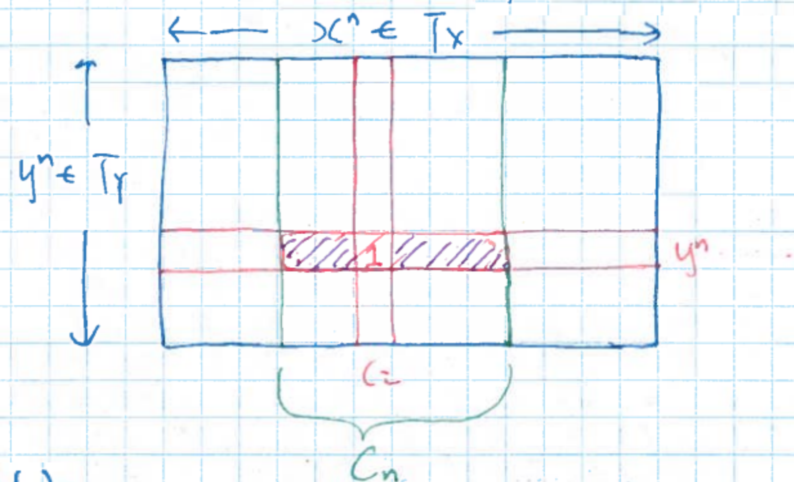
(a) By JAEP (1) with prob $\geq 1-\varepsilon$, $c_i y^n \in A_{n,\delta}$

(b) Assuming $c_i y^n \in A_{n,\delta}$:

error if $\exists$ $k \neq i$ but $c_k y^n \in A_{n,\delta}$

i.e., there is at least another "1" in the purple region, besides ci.

Joint typicality table: (x^n, y^n)-entry $= 1$ iff $x^n y^n \in A_{n,\delta}$



For any $k \neq i$, ck, ci independent, so ck, y^n also independent.

By JAEP (c),

$$\Pr(c_k y^n \in A_{n,\delta}) \leq 2^{-n(I(X;Y)-3\delta)}$$

over C_n, $N^{\otimes n}$

So, for this fixed i, there is an error if

$c1, y^n$ or $c2, y^n$ or $c3, y^n$ ... or c_{i-1} y^n or c_{i+1} y^n ... or c_M y^n

is in $A_{n,\delta}$.

$$\text{Prob of error} \leqslant \sum_{k \neq i} \text{Prob}(ck \text{ in } A_{n,\delta}) \qquad \text{(union bound)}$$

$$\leqslant \sum_{k \neq i} 2^{-n(I(X:Y)-3\delta)} \qquad \text{(previous page)}$$

$$\leqslant |M| \, 2^{-n(I(X:Y)-3\delta)}$$

Averaged over the choice of the code and channel noise,
assuming ci y^n jointly typical, and for any value of i.

Putting (a) and (b) together, for any i,

$$\mathbb{E}_{C_n} P_e(i) \leqslant \varepsilon + |M| \cdot 2^{-n(I(X:Y)-3\delta)}$$

Now, average over both the code and i:

$$\mathbb{E}_{C_n} \bar{P}_e(C_n) = \mathbb{E}_{C_n} \frac{1}{M} \sum_{i=1}^{M} P_e(i) \quad (\text{using } C_n)$$

$$= \frac{1}{M} \sum_{i=1}^{M} \mathbb{E}_{C_n} P_e(i)$$

$$\le \frac{1}{M} \sum_{i=1}^{M} \left(\delta + |M| \cdot 2^{-n(I(X:Y)-3\delta)} \right)$$

$$= \varepsilon + |M| \cdot 2^{-n(I(X:Y)-3\delta)}$$

Finally, choosing our parameters for part (1):

$$\therefore \forall \; |M| = 2^{nR}, \quad R < I(X:Y) - 3\delta - \frac{1}{n}\log\left(\frac{1}{\eta}\right)$$

$$\varepsilon < \eta$$

We have $\mathbb{E}_{C_n} \bar{P}_e(C_n) \le 2\eta$.

low error averaged over code, over i

at least one code has low error (averaged over i)

Part (2) follows immediately:

$$\therefore \exists \; \tilde{C}_n \text{ s.t. } \bar{P}_e(C_n) \le 2\eta.$$

Step (3):

From $\hat{C}_n$, we can get a code $\tilde{\tilde{C}}_n$ with $P_e(\tilde{\tilde{C}}_n) \le 2 \cdot \bar{P}_e(\hat{C}_n)$.

$\tilde{\tilde{C}}_n$ consists of $\frac{M}{2}$ codewords in $\hat{C}_n$ with the smallest probs of error.

~~with prob of error less than the median~~

(i.e. $\tilde{\tilde{C}}_n$ = "better half" of $\hat{C}_n$).

What is the worse error for the better half? 4η

Proof: If not, the best error for the worse half is more than 4η

and the average error (over codewords) will exceed 2η

What is the effect on the rate? $\tilde{\tilde{C}}_n$ sends 1 fewer bit than $\hat{C}_n$

Rate ↓ by $\frac{1}{n}$, which is $I(X;Y) - 3\eta - \frac{1}{n}\log\left(\frac{1}{\eta}\right) - \frac{1}{n}$

$\therefore$ $I(X;Y)$ is an achievable rate for N.

Finally, max over $p(x)$, $\max_{p(x)} I(X;Y)$ is achievable for N.

which completes the direct coding half of the capacity theorem.

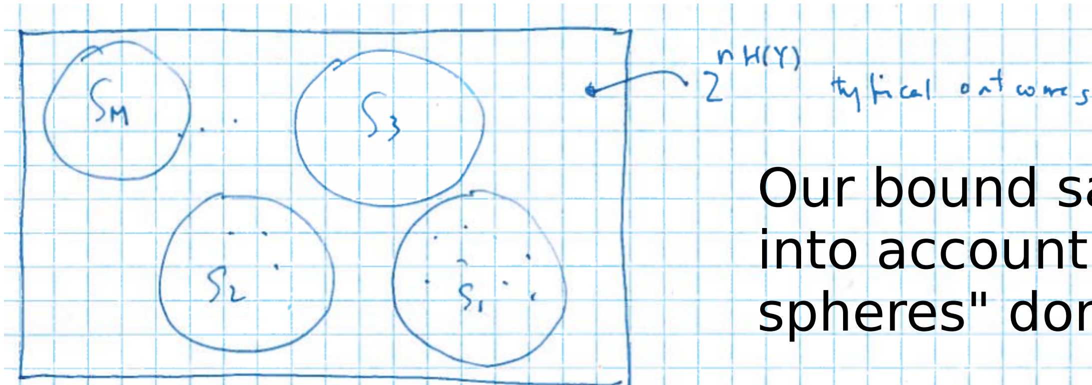
Remarks:

NB: Once C_n chosen $X_1 X_2 \ldots X_n$ <u>not</u> independent! That's why $\mathbb{E}_{C_n}$ is needed.

Techniques: Random codes, symmetry, existential proof

Expunging half of the worse codeword to boost $\overline{P_e}$ to P_e.

One more geometric picture (to complement the Hamming spheres):



Our bound says that, taking probabilities into account, these "inverse Hamming spheres" don't overlap much.

where $S_i = \{y^n : (c_i, y^n) \in A_{n,\delta}\}$, $|S_i| \approx 2^{nH(Y|X)}$.

Intuitively, fitting $\ll 2^{nH(Y)} / 2^{nH(Y|X)} \approx 2^{nI(X;Y)}$ spheres works.

The JAEP table says a lot about how these spheres are distributed.

Proving the capacity theorem

(1) Prove a <u>direct coding theorem</u>:

In this case, given any $p(x)$, show $\exists$ codes achieving the rate $I(X:Y)$ such that $P_e(C_n) \to 0$.

This gives $C(N) \geq \max_{p(x)} I(X:Y)$.

(2) Prove a <u>converse</u>:

For any achievable R, $R \leq \max_{p(x)} I(X:Y)$,

which gives $C(N) \leq \max_{p(x)} I(X:Y)$.

Here, the upper bound (converse) on the capacity matches the lower bound achieved by codes -- so we know the capacity expression.

Some terminologies:

Statement: A implies B

$\Leftrightarrow$

Converse of statement
B implies A

Contrapositive of statement
not-B implies not-A

Direct coding theorem: if $R \le \max_{p(x)} I(X{:}Y)$ then R achievable

Converse to the direct coding theorem:

if R achievable, then $R \le \max_{p(x)} I(X{:}Y)$

Contrapositive: (of the converse)

If $R > \max_{p(x)} I(X{:}Y)$, any sequence of $(\lceil 2^{nR} \rceil, n)$ codes has $P_e(C_n) \not\to 0$.

Strong converse: $1 - P_e(C(n)) \sim \exp(-n \times \text{const})$.

Proof of converse: if R achievable, then $R \leq \max\limits_{p(x)} I(X:Y)$

Let R be achievable

Let C_n be the sequence of $(\lfloor 2^{nR} \rfloor, n)$ codes with $P_e(C_n) \to 0$

Let W_n be a rv describing a random element in $\{1, 2, \dots, \lfloor 2^{nR} \rfloor\}$

(the index set for messages from C_n).

Let $p_n = P_e(C_n)$

The following holds from definition:

$$[H(W_n, Y^n) - H(Y^n)] + [H(W_n) + H(Y^n) - H(W_n, Y^n)]$$

$$nR = H(W_n) = H(W_n \mid Y^n) + I(W_n : Y^n)$$

We will prove
3 inequalities:

① $\leq$

$1 + p_n nR$

② $\leq$

$I(\mathcal{E}_n(W_n) : Y^n)$

③ $\leq$

$n \max\limits_{p(x)} I(X:Y)$

The following holds from definition:

$$\overbrace{nR = H(W_n) = H(W_n \mid Y^n)}^{[H(W_n, Y^n) - H(Y^n)]} + \overbrace{I(W_n : Y^n)}^{[H(W_n) + H(Y^n) - H(W_n, Y^n)]}$$

$$\text{①} \quad H(W_n \mid Y^n) \leq 1 + p_n\, nR$$

$$\text{②} \quad I(W_n : Y^n) \leq I(\mathcal{E}_n(W_n) : Y^n)$$

$$\text{③} \quad I(\mathcal{E}_n(W_n) : Y^n) \leq n \max_{p(x)} I(X : Y)$$

Together $nR \leq 1 + p_n\, nR + n \max_{p(x)} I(X:Y)$

Divide by n, take $n \to \infty$, $p_n \to 0$, we have

$$R \leq \max_{p(x)} I(X:Y).$$

We now prove the first of these 3 inequalities:

Thm [Fanos ineq]

Consider r.v.s A, B, C, $C = f(B)$, f function

Let $q = \text{prob}(A \neq C)$

Ω = sample space of A

Then $h(q) + q \log(|\Omega| - 1) \geq H(A|B)$.

Pf: Define new rv E s.t $E = \begin{cases} 0 & \text{if } A = C \\ 1 & \text{otherwise} \end{cases}$

$$\underbrace{H(EA|B)}_{H(EAB)-H(B)} = \underbrace{H(A|B)}_{H(AB)-H(B)} + \underbrace{H(\cancel{E}|AB)}_{H(EAB)-H(AB)} \qquad = H(E|B) + H(A|EB)$$

exchange E & A

$$\therefore \; H(A|B) + 0 = H(E|B) + H(A|EB)$$

$$\leq H(E) + \sum_b p(b) \left[q\, H(A|E=1\; B=b) + (1-q) H(A|\cancel{E=0\; B=b}) \right]$$

$$\therefore \; H(A|B) \leq h(q) + q\, H(A|E=1\; B)$$

$$\leq h(q) + q \log\left[|\Omega| - 1\right]$$

We have used:

(H8) Conditioning reduces entropy

$H(X|Y) \leq H(X)$, "=" iff X, Y indep.

Def [Conditional entropy]

Using the above notations, the entropy of X conditioned on Y is:

$$H(X|Y) := \sum_y p(y) \underbrace{H(q_y)}_{\substack{\text{entropy of } X \\ \text{given } Y=y}}$$

(H1) Range: $0 \leq H(X) \leq \log |\Omega|$

"=" iff $\exists a$ s.t $p(x)=0 \ \forall x \neq a$

"=" iff $p(x) = \frac{1}{|\Omega|} \ \forall x$

We now prove the first of these 3 inequalities:

Thm [Fanos ineq]

Consider r.v.s A, B, C, $C = f(B)$, f function

Let q = prob $(A \neq C)$

Ω = sample space of A

Then $h(q) + q \log(|\Omega| - 1) \geq H(A|B)$.

• Set $A = W_n$, $B = Y^n$, $|\Omega| = 2^{nR}$, $f = D_n$, $q = p_n$

Then $H(W_n | Y^n) \leq h(p_n) + p_n \cdot nR$.

$\leq 1 + p_n \cdot nR$.

output

error prob

what is the input (from 1, 2, ..., $2^{\lfloor nR \rfloor}$)

We now prove the second of these 3 inequalities:

$$I(W_n : Y^n) \overset{②}{\geq} I(\mathcal{E}_n(W_n) : Y^n)$$

From H11, if A->B->C is a Markov chain, then, I(A:B) >= I(A:C).

Note also from the proof of H11 that A->B->C is a Markov chain iff I(A:C|B)=0 iff C->B->A is a Markov chain, so, I(C:B) >= I(C:A).

Set $A = W_n$, $B = \mathcal{E}_n(W_n)$, $C = Y^n$

So $I(\mathcal{E}_n(W_n) : Y^n) \geq I(W_n : Y^n)$

We now prove the third of these 3 inequalities:

$$I(\mathcal{E}_n(W_n) : Y^n) \overset{③}{\leq} n \max_{p(x)} I(X : Y)$$

Lemma: Let $Y^n = N^{\otimes n}(X^n)$

Then $I(X^n : Y^n) \leq \sum_{i=1}^{n} I(X_i : Y_i)$

NB: neither X^n nor Y^n need to be iid

cf. X^n from codewords, $C_i = \underbrace{x_{i1} x_{i2} \cdots x_{in}}_{\text{correlated}}$

We now prove the third of these 3 inequalities:

$$I(\mathcal{E}_n(W_n) : Y^n) \overset{③}{\leq} n \max_{p(x)} I(X:Y)$$

Lemma: Let $Y^n = N^{\otimes n}(X^n)$

Then $I(X^n : Y^n) \leq \sum_{i=1}^{n} I(X_i : Y_i)$

Pf: $I(X^n : Y^n) = H(Y^n) - H(Y^n | X^n)$

$= H(Y^n) - \sum_{i=1}^{n} H(Y_i | Y_1 Y_2 \ldots Y_{i-1} X^n)$ chain rule with cond.

$= H(Y^n) - \sum_{i=1}^{n} H(Y_i | X_i)$ with X_i, Y_i does not depend on the rest.

$\leq \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i | X_i)$ subadditivity

$= \sum_{i=1}^{n} I(X_i : Y_i)$.

We now prove the third of these 3 inequalities:

$$I(\mathcal{E}_n(W_n):Y^n) \overset{(3)}{\leq} n \max_{p(x)} I(X:Y)$$

Lemma: Let $Y^n = N^{\otimes n}(X^n)$

Then $I(X^n:Y^n) \leq \sum_{i=1}^{n} I(X_i:Y_i)$

To get the third inequality, note $X^n = \mathcal{E}_n(W_n)$,

$$I(\mathcal{E}_n(W_n):Y^n) = I(X^n:Y^n) \leq \sum_{i=1}^{n} I(X_i:Y_i)$$ from lemma

$$\leq n \max_{p(x)} I(X:Y)$$

This completes the proof of the converse, and also the capacity theorem.

NB. Back classical communication from Bob to Alice does not affect the proof of the converse, so, the same upper bound for the rate holds; compared to the direct coding theorem WITHOUT the back comm, it shows that classical "feedback" does not increase capacity (though it may reduce code complexity etc). e.g., erasure channel.