COMP2610/6261 - Information Theory

Lecture 19: Block Codes and the Coding Theorem

Robert C. Williamson

Research School of Computer Science

Australian National University

9 October, 2018

Channel Capacity: Recap

The *largest possible* reduction in uncertainty achievable across a channel is its **capacity**

Channel Capacity

The capacity C of a channel Q is the largest mutual information between its input and output for any choice of input ensemble. That is,

$$C = \max_{\mathbf{p}_X} I(X; Y)$$

1. Block Codes
2. The Noisy Channel Coding Theorem
3. Extended Channels

Communicating over Noisy Channels

Suppose we know we have to communicate over some channel Q and we want build an *encoder/decoder* pair to reliably send a message **s** over Q.

$$\boxed{\mathbf{s}_{in}} \xrightarrow{\text{Encoder}} \boxed{\mathbf{x}} \xrightarrow{Q} \boxed{\mathbf{y}} \xrightarrow{\text{Decoder}} \boxed{\mathbf{s}_{out}}$$

Block Codes

We now consider codes that make repeated use of a noisy channel to communicate a predefined set of messages $\mathcal{S} = \{1, 2, \ldots, S\}$

Block Codes

We now consider codes that make repeated use of a noisy channel to communicate a predefined set of messages $\mathcal{S} = \{1, 2, \ldots, S\}$

Recall a general encoder is of the form

$$\text{enc} : \mathcal{S} \to \mathcal{X}^N$$

Equivalently, each $s \in \mathcal{S} = \{1, 2, \ldots, S\}$ is paired with a unique *block* of symbols $\mathbf{x} \in \mathcal{X}^N$

Block Codes

We now consider codes that make repeated use of a noisy channel to communicate a predefined set of messages $\mathcal{S} = \{1, 2, \ldots, S\}$

Recall a general encoder is of the form

$$\text{enc}: \mathcal{S} \to \mathcal{X}^N$$

Equivalently, each $s \in \mathcal{S} = \{1, 2, \ldots, S\}$ is paired with a unique *block* of symbols $\mathbf{x} \in \mathcal{X}^N$

Thus, we can imagine there being S unique codewords $\{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(S)}\}$, where each codeword has block length N

Block Codes: Example

Suppose $\mathcal{S} = \{1, 2, 3, 4\}$

Message ID s	**Message encoding**
1	$\texttt{00}$
2	$\texttt{01}$
3	$\texttt{10}$
4	$\texttt{11}$

Block size $N = 2$

Codewords $\mathbf{x}^{(1)} = \texttt{00}, \mathbf{x}^{(2)} = \texttt{01}$, and so on

Block Codes: Formally

We formalise the preceding with the following notion:

(N, K) Block Code

Given a channel Q with inputs $\mathcal{X}$ and outputs $\mathcal{Y}$, an integer $N > 0$, and $K > 0$, an (N, K) Block Code for Q is a list of $S = 2^K$ codewords

$$\mathcal{C} = \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(2^K)}\}$$

where each $\mathbf{x}^{(s)} \in \mathcal{X}^N$ consists of N symbols from $\mathcal{X}$.

The code is parameterised by the length of the block, and the number of messages that are encoded

- We parametrise by $K = \log_2 S$ for mathematical convenience
- Doesn't have to be an integer

Block Codes and Rates

An (N, K) block code makes N uses of a channel to transmit one of S possible outcomes

We can measure the amount of 'information' contained in each use as:

Rate of an (N, K) Block Code

The **rate** of an (N, K) block code is $\frac{\log_2 S}{N} = \frac{K}{N}$ bits per channel use.

Block Codes: Examples

Examples (for Binary Symmetric Channel Q)

- A $(1, 1)$ block code: $\mathcal{C} = \{0, 1\}$ — Rate: 1

- A $(3, 2)$ block code: $\mathcal{C} = \{000, 001, 100, 111\}$ — Rate: $\frac{2}{3}$

- A $(3, \log_2 3)$ block code: $\mathcal{C} = \{001, 010, 100\}$ — Rate: $\frac{\log_2 3}{3} \approx 0.53$

Decoding Block Codes

An (N, K) block code sends each message $s \in \mathcal{S} = \{1, 2, \ldots, 2^K\}$ over a channel Q as $\mathbf{x}^s \in \mathcal{X}^N$

The receiver sees the block $\mathbf{y} \in \mathcal{Y}^N$, and attempts to infer s via some

$$\texttt{dec}\colon \mathcal{Y}^N \to \mathcal{S}$$

Decoding Block Codes

An (N, K) block code sends each message $s \in \mathcal{S} = \{1, 2, \ldots, 2^K\}$ over a channel Q as $\mathbf{x}^s \in \mathcal{X}^N$

The receiver sees the block $\mathbf{y} \in \mathcal{Y}^N$, and attempts to infer s via some

$$\texttt{dec} \colon \mathcal{Y}^N \to \mathcal{S}$$

Even if $\mathcal{X} = \mathcal{Y}$, the decoder must allow for any $\mathcal{Y}^N$, not just the expected codewords $\{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(2^K)}\}$

Decoding Block Codes: Formally

Block Decoder

A **decoder** for a (N, K) block code is a mapping that associates each $\mathbf{y} \in \mathcal{Y}^N$ with an $\hat{s} \in \{1, 2, \ldots, 2^K\}$.

Decoding Block Codes: Formally

Block Decoder

A **decoder** for a (N, K) block code is a mapping that associates each $\mathbf{y} \in \mathcal{Y}^N$ with an $\hat{s} \in \{1, 2, \ldots, 2^K\}$.

Ideally, we would like the decoded $\hat{s}$ to be maximally likely to be equal to s

Optimal Decoder

An **optimal decoder** for a code $\mathcal{S}$, channel Q, and *prior* $P(s)$ maps $\mathbf{y}$ to $\hat{s}$ such that $P(\hat{s}|\mathbf{y})$ is maximal.

That is, $\mathtt{dec}_{opt}(\mathbf{y}) = \arg\max_s P(s|\mathbf{y}) = \arg\max_s P(\mathbf{y}|s) \cdot P(s)$

Decoding Block Codes: Examples

Example The $(2,1)$ block code $\mathcal{S} = \{000, 111\}$ and majority vote decoder $d : \{0,1\}^3 \to \{1,2\}$ defined by

$$d(000) = d(001) = d(010) = d(100) = 1$$
$$d(111) = d(110) = d(101) = d(011) = 2$$

Rates and Reliability

Ideally, we would like to have high rate for our channel code

- Low rate implies that we are being "wasteful" with our channel use

But intuitively, at high rates, we run the risk of losing reliability

- If N is small, we may be more easily "confused" about an input

How to measure reliability?

Reliability

Want an *encoder/decoder* pair to reliably send a messages over channel Q.

$$\boxed{\mathbf{s}_{in}} \xrightarrow{\text{Encoder}} \boxed{\mathbf{x}} \xrightarrow{Q} \boxed{\mathbf{y}} \xrightarrow{\text{Decoder}} \boxed{\mathbf{s}_{out}}$$

Reliability

Want an *encoder/decoder* pair to reliably send a messages over channel Q.

$$\mathbf{s}_{in} \xrightarrow{\text{Encoder}} \mathbf{x} \xrightarrow{Q} \mathbf{y} \xrightarrow{\text{Decoder}} \mathbf{s}_{out}$$

Probability of (Block) Error

Given a channel Q the **probability of (block) error** for a code is

$$p_B = P(\mathbf{s}_{out} \neq \mathbf{s}_{in}) = \sum_{\mathbf{s}_{in}} P(\mathbf{s}_{out} \neq \mathbf{s}_{in} | \mathbf{s}_{in}) P(\mathbf{s}_{in})$$

and its **maximum probability of (block) error** is

$$p_{BM} = \max_{\mathbf{s}_{in}} P(\mathbf{s}_{out} \neq \mathbf{s}_{in} | \mathbf{s}_{in})$$

Reliability

Want an *encoder/decoder* pair to reliably send a messages over channel Q.

$$\boxed{\mathbf{s}_{in}} \xrightarrow{\text{Encoder}} \boxed{\mathbf{x}} \xrightarrow{Q} \boxed{\mathbf{y}} \xrightarrow{\text{Decoder}} \boxed{\mathbf{s}_{out}}$$

Probability of (Block) Error

Given a channel Q the **probability of (block) error** for a code is

$$p_B = P(\mathbf{s}_{out} \neq \mathbf{s}_{in}) = \sum_{\mathbf{s}_{in}} P(\mathbf{s}_{out} \neq \mathbf{s}_{in} | \mathbf{s}_{in}) P(\mathbf{s}_{in})$$

and its **maximum probability of (block) error** is

$$p_{BM} = \max_{\mathbf{s}_{in}} P(\mathbf{s}_{out} \neq \mathbf{s}_{in} | \mathbf{s}_{in})$$

As $P(\mathbf{s}_{out} \neq \mathbf{s}_{in} | \mathbf{s}_{in}) \leq p_{BM}$ for all $\mathbf{s}_{in}$ we get $p_B \leq \sum_{\mathbf{s}_{in}} p_{BM} P(\mathbf{s}_{in}) = p_{BM}$ and so if $p_{BM} \to 0$ then $p_B \to 0$.

Reliability: Example

Suppose $\mathbf{s} \in \{\mathtt{a}, \mathtt{b}\}$ and we encode by $\mathtt{a} \rightarrow \mathtt{000}$ and $\mathtt{b} \rightarrow \mathtt{111}$.
To decode we count the number of 1s and 0s and set all bits to the majority count to determine $\mathbf{s}$

$$\underbrace{\mathtt{000}, \mathtt{001}, \mathtt{010}, \mathtt{100}}_{A} \rightarrow \mathtt{a} \quad \text{and} \quad \underbrace{\mathtt{111}, \mathtt{110}, \mathtt{101}, \mathtt{011}}_{B} \rightarrow \mathtt{b}$$

Reliability: Example

Suppose $\mathbf{s} \in \{\mathtt{a}, \mathtt{b}\}$ and we encode by $\mathtt{a} \to \mathtt{000}$ and $\mathtt{b} \to \mathtt{111}$.
To decode we count the number of 1s and 0s and set all bits to the majority count to determine $\mathbf{s}$

$$\underbrace{\mathtt{000}, \mathtt{001}, \mathtt{010}, \mathtt{100}}_{A} \to \mathtt{a} \quad \text{and} \quad \underbrace{\mathtt{111}, \mathtt{110}, \mathtt{101}, \mathtt{011}}_{B} \to \mathtt{b}$$

If the channel Q is binary symmetric,

$$\begin{aligned} p_B &= P(\mathbf{s}_{in} \neq \mathbf{s}_{out}) \\ &= P(\mathbf{y} \in B | 000)\, p_{\mathrm{a}} + P(\mathbf{y} \in A | 111)\, p_{\mathrm{b}} \\ &= [f^3 + 3f^2(1-f)]\, p_{\mathrm{a}} + [f^3 + 3f^2(1-f)]\, p_{\mathrm{b}} \\ &= f^3 + 3f^2(1-f). \end{aligned}$$

In fact,

$$p_{BM} = \max(P(\mathbf{y} \in B | 000), P(\mathbf{y} \in A | 111)) = f^3 + 3f^2(1-f).$$

Achievable Rates

Ideally, we would like to consider rates of transmission for which we can guarantee small maximum probability of block error

Even more ideally, we would like rates for which we can guarantee arbitrarily small maximum probability of block error

- We will call such rates achievable

Achievable Rates

Ideally, we would like to consider rates of transmission for which we can guarantee small maximum probability of block error

Even more ideally, we would like rates for which we can guarantee arbitrarily small maximum probability of block error

- We will call such rates achievable

Achievable Rate

A rate R over a channel Q is said to be **achievable** if, for any $\epsilon > 0$ there is a (N, K) block code and decoder such that its rate $K/N \geq R$ and its maximum probability of block error satisfies

$$p_{BM} = \max_{\mathbf{s}_{in}} P(\mathbf{s}_{out} \neq \mathbf{s}_{in} | \mathbf{s}_{in}) < \epsilon$$

Achievable Rates

Achievable rates sound nice in theory, but surely they cannot exist?

- Surely we will have to drive $R \to 0$ to get small error probability?

Achievable Rates

Achievable rates sound nice in theory, but surely they cannot exist?

- Surely we will have to drive $R \to 0$ to get small error probability?

Remarkably, we have:

Noisy-Channel Coding Theorem (Brief)

If Q is a channel with capacity C then the rate R is *achievable* **if and only if** $R \leq C$, that is, the rate is no greater than the channel capacity.

The Noisy-Channel Coding Theorem

Example

Example:

- In last lecture: BSC Q with $f = 0.15$ has capacity $C = 0.39$ bits.

- Suppose we want error less than $\epsilon = 0.05$ and rate $R > 0.25$

- The NCCT tells us there should be, for N large enough, an (N, K) code with $K/N \geq 0.25$

Indeed, we showed the code $\mathcal{S} = \{000, 111\}$ with majority vote decoder has probability of error $0.028 < 0.05$ for Q and rate $1/3 > 0.25$.

Example:

- In last lecture: BSC Q with $f = 0.15$ has capacity $C = 0.39$ bits.

- Suppose we want error less than $\epsilon = 0.05$ and rate $R > 0.25$

- The NCCT tells us there should be, for N large enough, an (N, K) code with $K/N \geq 0.25$

Indeed, we showed the code $\mathcal{S} = \{000, 111\}$ with majority vote decoder has probability of error $0.028 < 0.05$ for Q and rate $1/3 > 0.25$.

- For $N = 3$ there is a $(3, 1)$ code meeting the requirements.

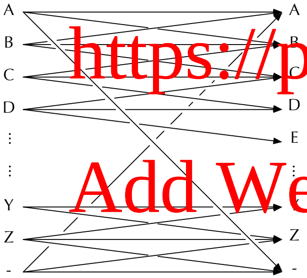
- However, there is *no code* with same ϵ and rate $1/2 > 0.39 = C$.

The Noisy Typewriter Channel

This channel simulates a noisy "typewriter". Inputs and outputs are 26 letters A through Z plus space. With probability $\frac{1}{3}$, each letter is either: unchanged; changed to the next letter; changed to the previous letter.

The Noisy Typewriter Channel

This channel simulates a noisy “typewriter”. Inputs and outputs are 26 letters A through Z plus space. With probability $\frac{1}{3}$, each letter is either: unchanged; changed to the next letter; changed to the previous letter.



Inputs $\mathcal{X} = \{\mathtt{A}, \mathtt{B}, \ldots, \mathtt{Z}, \_\}$;

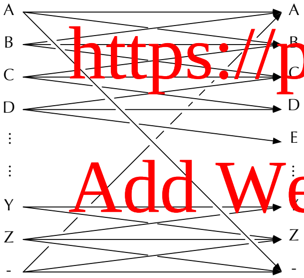
Outputs $\mathcal{Y} = \{\mathtt{A}, \mathtt{B}, \ldots, \mathtt{Z}, \_\}$;

Transition probabilities

$$Q = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & 0 & 0 & \cdots & 0 & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & \cdots & 0 & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{1}{3} & 0 & 0 & & \cdots & \frac{1}{3} & \frac{1}{3} \end{bmatrix}$$

The Noisy Typewriter Channel

This channel simulates a noisy "typewriter". Inputs and outputs are 26 letters A through Z plus space. With probability $\frac{1}{3}$, each letter is either: unchanged; changed to the next letter; changed to the previous letter.



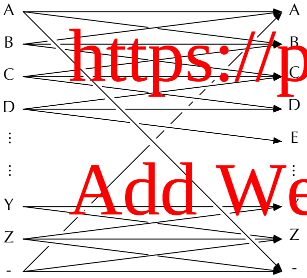
Inputs $\mathcal{X} = \{\mathtt{A}, \mathtt{B}, \ldots, \mathtt{Z}, \_\}$;
Outputs $\mathcal{Y} = \{\mathtt{A}, \mathtt{B}, \ldots, \mathtt{Z}, \_\}$;
Transition probabilities

$$Q = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & 0 & 0 & \cdots & 0 & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & \cdots & 0 & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{1}{3} & 0 & 0 & & \cdots & \frac{1}{3} & \frac{1}{3} \end{bmatrix}$$

The transition matrix for this channel has a diagonal structure: all of the probability mass is concentrated around the diagonal.

The Noisy Typewriter Channel

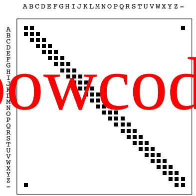
This channel simulates a noisy "typewriter". Inputs and outputs are 26 letters A through Z plus space. With probability $\frac{1}{3}$, each letter is either: unchanged; changed to the next letter, changed to the previous letter.



Inputs $\mathcal{X} = \{\mathtt{A}, \mathtt{B}, \ldots, \mathtt{Z}, \text{\_}\}$;

Outputs $\mathcal{Y} = \{\mathtt{A}, \mathtt{B}, \ldots, \mathtt{Z}, \text{\_}\}$;

Transition probabilities



The transition matrix for this channel has a diagonal structure: all of the probability mass is concentrated around the diagonal.

Noisy Channel Coding Theorem

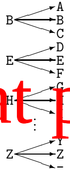
NCCT

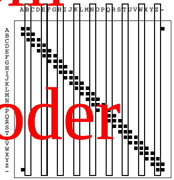
Any rate $R < C$ is *achievable* for Q (i.e., for any tolerance $\epsilon > 0$, an (N, K) code with rate $K/N \geq R$ exists with max. block error $p_{BM} < \epsilon$)

Consider a simple example:

For noisy typewriter Q:

- The capacity is $C = \log_2 9$
- For any $\epsilon > 0$ and $R < C$ we can choose $N = 1$ . . .
- . . . and code messages using $\mathcal{C} = \{\mathtt{B}, \mathtt{E}, \ldots, \mathtt{Z}\}$





Noisy Channel Coding Theorem

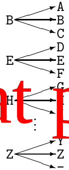
NCCT

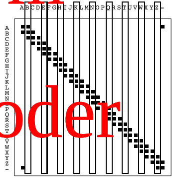
Any rate $R < C$ is *achievable* for Q (i.e., for any tolerance $\epsilon > 0$, an (N, K) code with rate $K/N \geq R$ exists with max. block error $p_{BM} < \epsilon$)

Consider a simple example:

For noisy typewriter Q:

- The capacity is $C = \log_2 9$
- For any $\epsilon > 0$ and $R < C$ we can choose $N = 1$ . . .
- . . . and code messages using $\mathcal{C} = \{\mathtt{B}, \mathtt{E}, \ldots, \mathtt{Z}\}$





Since $|\mathcal{C}| = 9$ we have $K = \log_2 9$ so $K/N = \log_2 9 \geq R$ for any $R < C$, and $\mathcal{C}$ has zero error so $p_{BM} = 0 < \epsilon$

Noisy Channel Coding Theorem: How Is This Possible?

The main "trick" to minimising p_{BM} is to construct a (N, K) block code with (almost) **non-confusable** codes

- A code such that the set of $\mathbf{y}$ that each $\mathbf{x}^{(s)}$ are sent to by Q have low probability intersection

Noisy Channel Coding Theorem: How Is This Possible?

The main "trick" to minimising p_{BM} is to construct a (N, K) block code with (almost) **non-confusable** codes

- A code such that the set of $\mathbf{y}$ that each $\mathbf{x}^{(s)}$ are sent to by Q have low probability intersection

This is possible because extended channels look like the noisy typewriter!

Extended Channels

When used N times, a channel Q from $\mathcal{X}$ to $\mathcal{Y}$ can be seen as an *extended channel* taking "symbols" from $\mathcal{X}^N$ to "symbols" in $\mathcal{Y}^N$.

Extended Channel

The N^{th} **extended channel** of Q from $\mathcal{X}$ to $\mathcal{Y}$ is a channel from $\mathcal{X}^N$ to $\mathcal{Y}^N$ with transition probability from $\mathbf{x} \in \mathcal{X}^N$ to $\mathbf{y} \in \mathcal{Y}^N$ given by

$$P(\mathbf{y}|\mathbf{x}) = \prod_{n=1}^{N} P(y_i|x_i)$$

Extended Channels

When used N times, a channel Q from $\mathcal{X}$ to $\mathcal{Y}$ can be seen as an *extended channel* taking "symbols" from $\mathcal{X}^N$ to "symbols" in $\mathcal{Y}^N$.

Extended Channel

The N^{th} **extended channel** of Q from $\mathcal{X}$ to $\mathcal{Y}$ is a channel from $\mathcal{X}^N$ to $\mathcal{Y}^N$ with transition probability from $\mathbf{x} \in \mathcal{X}^N$ to $\mathbf{y} \in \mathcal{Y}^N$ given by

$$P(\mathbf{y}|\mathbf{x}) = \prod_{n=1}^{N} P(y_i|x_i)$$

Example: *BSC* Q with $f = 0.1$ from $\mathcal{X} = \{0, 1\}$ to $\mathcal{Y} = \{0, 1\}$ has $N = 2$ *extended channel* from $\mathcal{X}^2 = \{00, 01, 10, 11\}$ to $\mathcal{Y}^2 = \{00, 01, 10, 11\}$ with

$$Q_2 = \begin{bmatrix} 0.81 & 0.09 & 0.09 & 0.01 \\ 0.09 & 0.81 & 0.01 & 0.09 \\ 0.09 & 0.01 & 0.81 & 0.09 \\ 0.01 & 0.09 & 0.09 & 0.81 \end{bmatrix}$$

Extended Channels

When used N times, a channel Q from $\mathcal{X}$ to $\mathcal{Y}$ can be seen as an *extended channel* taking "symbols" from $\mathcal{X}^N$ to "symbols" in $\mathcal{Y}^N$.

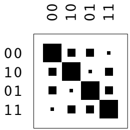
Extended Channel

The N^{th} **extended channel** of Q from $\mathcal{X}$ to $\mathcal{Y}$ is a channel from $\mathcal{X}^N$ to $\mathcal{Y}^N$ with transition probability from $\mathbf{x} \in \mathcal{X}^N$ to $\mathbf{y} \in \mathcal{Y}^N$ given by

$$P(\mathbf{y}|\mathbf{x}) = \prod_{n=1}^{N} P(y_i|x_i)$$

Example: BSC Q with $f = 0.1$ from $\mathcal{X} = \{0, 1\}$ to $\mathcal{Y} = \{0, 1\}$ has $N = 2$ *extended channel* from $\mathcal{X}^2 = \{00, 01, 10, 11\}$ to $\mathcal{Y}^2 = \{00, 01, 10, 11\}$ with

$$Q_2 = \begin{bmatrix} 0.81 & 0.09 & 0.09 & 0.01 \\ 0.09 & 0.81 & 0.01 & 0.09 \\ 0.09 & 0.01 & 0.81 & 0.09 \\ 0.01 & 0.09 & 0.09 & 0.81 \end{bmatrix}$$

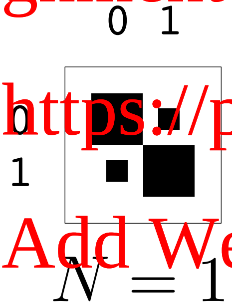


Extended Channels and the Noisy Typewriter

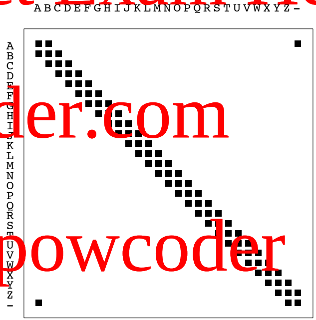
As N increases, any extended channel looks like the noisy typewriter!

Extended Channels and the Noisy Typewriter

As N increases, any extended channel looks like the noisy typewriter!



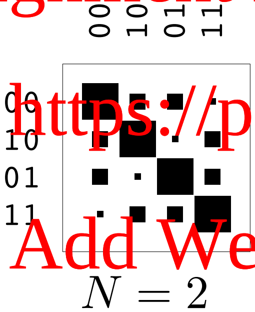
Extended Binary Symmetric Channel



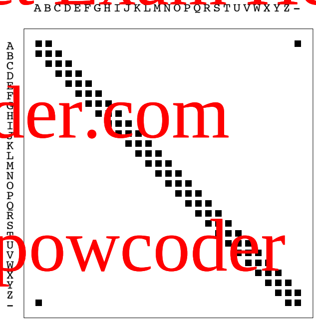
Noisy Typewriter Channel

Extended Channels and the Noisy Typewriter

As N increases, any extended channel looks like the noisy typewriter!



Extended Binary Symmetric Channel



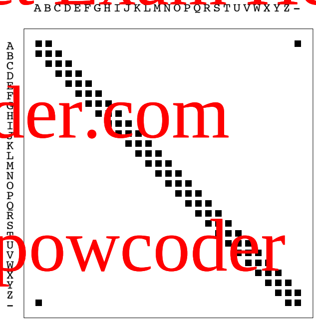
Noisy Typewriter Channel

Extended Channels and the Noisy Typewriter

As N increases, any extended channel looks like the noisy typewriter!

$N = 4$

Extended Binary Symmetric Channel

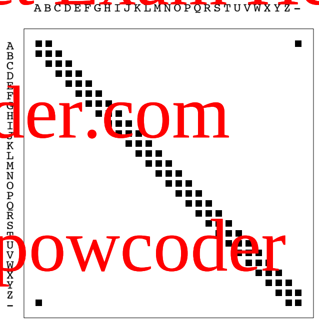


Noisy Typewriter Channel

Extended Channels and the Noisy Typewriter

As N increases, any extended channel looks like the noisy typewriter!

Extended Z Channel

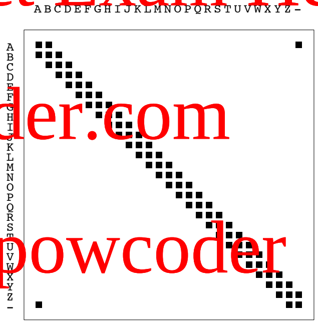


Noisy Typewriter Channel

Extended Channels and the Noisy Typewriter

As N increases, any extended channel looks like the noisy typewriter!

Extended Z Channel



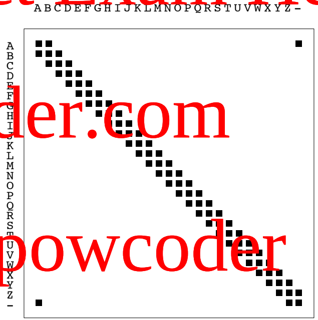
Noisy Typewriter Channel

Extended Channels and the Noisy Typewriter

As N increases, any extended channel looks like the noisy typewriter!

$N = 4$

Extended Z Channel



Noisy Typewriter Channel

Extended Channels and the Noisy Typewriter

Why does this happen?

Remember that as N gets larger, sequences $\mathbf{x} = x_1 x_2 \ldots x_N$ start looking typical

For a given $\mathbf{x}$, the corresponding $p(\mathbf{y} \mid \mathbf{x})$ will also be concentrated on a few sequences

Formalising this will require a notion of joint typicality

Summary and Reading

Main Points

- The Noisy Typewriter
- Extended Channels
- Block Codes
- The Noisy-Channel Coding Theorem (Statement only)

Reading

- MacKay §9.6
- Cover & Thomas §7.5

Summary and Reading

Main Points

- The Noisy Typewriter

- Extended Channels
- Block Codes

- The Noisy-Channel Coding Theorem (Statement only)

Reading

- MacKay §9.6

- Cover & Thomas §7.5

Next time: Detail of the NCCT, joint typicality, and a sketch of the proof!