COMP2610/6261 - Information Theory

Lecture 19: Block Codes and the Coding Theorem

Mark Reid and Aditya Menon

Research School of Computer Science
The Australian National University

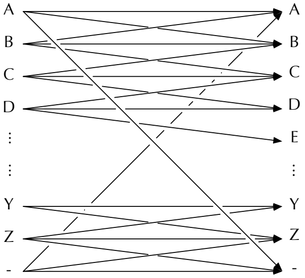
October 7th, 2014

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.

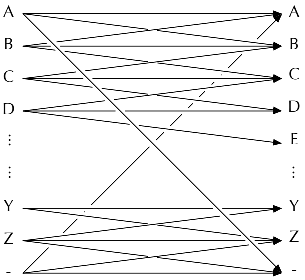


Inputs $\mathcal{X} = \{\mathtt{A}, \mathtt{B}, \ldots, \mathtt{Z}, \_\}$;
Outputs $\mathcal{Y} = \{\mathtt{A}, \mathtt{B}, \ldots, \mathtt{Z}, \_\}$;
Transition probabilities

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

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.



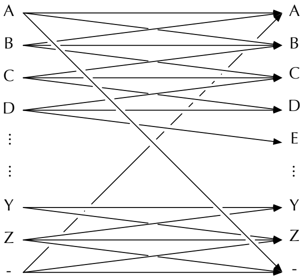
Inputs $\mathcal{X} = \{\texttt{A}, \texttt{B}, \ldots, \texttt{Z}, \_\}$;
Outputs $\mathcal{Y} = \{\texttt{A}, \texttt{B}, \ldots, \texttt{Z}, \_\}$;
Transition probabilities

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

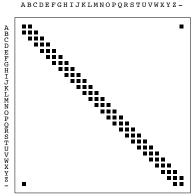
The transition matrix for this channel has a diagonal structure: all of the probability mass is concentrated around the diagonal.

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



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

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 $\mathcal{Y}^N$.

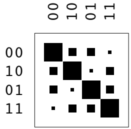
Extended Channel

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

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

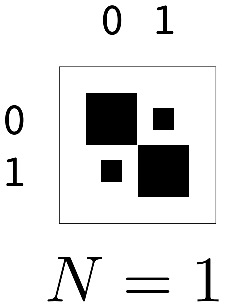


Extended Channels and the Noisy Typewriter

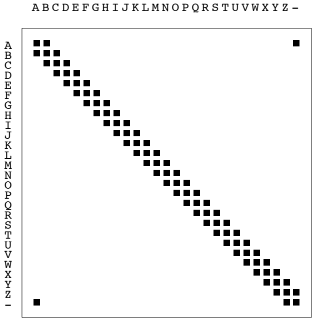
As N increases, any extended channel looks like the noisy typewriter!

Extended Channels and the Noisy Typewriter

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



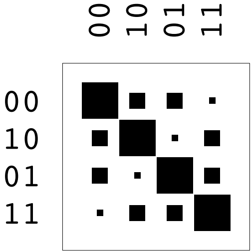
Extended Binary Symmetric Channel



Noisy Typewriter Channel

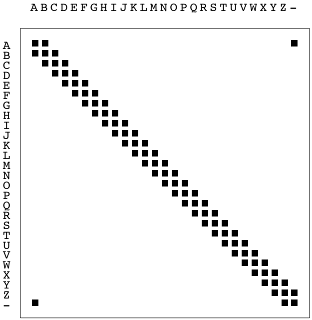
Extended Channels and the Noisy Typewriter

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



$N = 2$

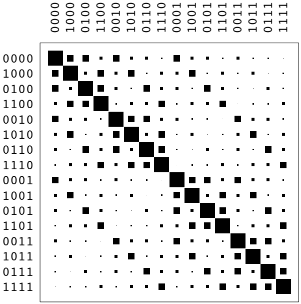
Extended Binary Symmetric Channel



Noisy Typewriter Channel

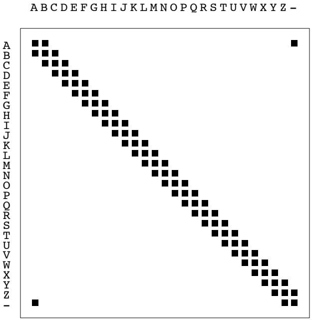
Extended Channels and the Noisy Typewriter

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



$N = 4$

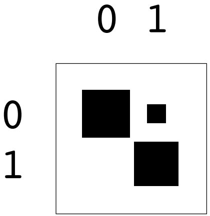
Extended Binary Symmetric Channel



Noisy Typewriter Channel

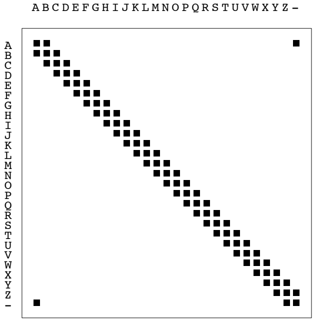
Extended Channels and the Noisy Typewriter

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



$$N = 1$$

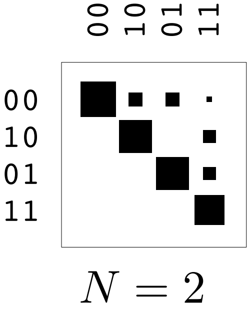
Extended Z Channel



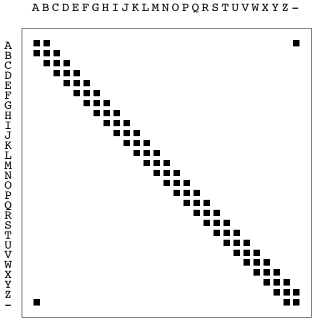
Noisy Typewriter Channel

Extended Channels and the Noisy Typewriter

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



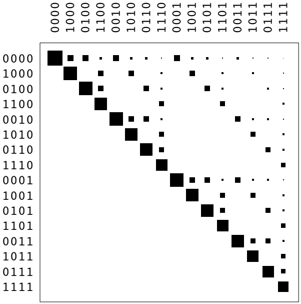
Extended Z Channel



Noisy Typewriter Channel

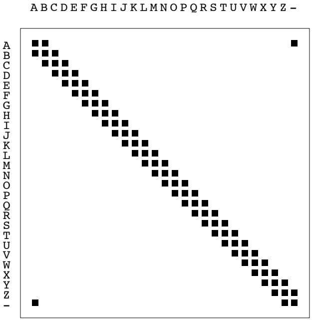
Extended Channels and the Noisy Typewriter

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



$N = 4$

Extended Z Channel



Noisy Typewriter Channel

Block Codes

We now formalise codes that make repeated use of a noisy channel to communicate a predefined set of S messages.

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

Block Codes

We now formalise codes that make repeated use of a noisy channel to communicate a predefined set of S messages.
Each $s \in \{1, 2, \ldots, S\}$ is paired with a unique *block* of symbols $\mathbf{x} \in \mathcal{X}^N$.

(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

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

where each $\mathbf{x}^{(s)} \in \mathcal{X}^N$ consists of N symbols from $\mathcal{X}$. The **rate** of such a block code is K/N bits per channel use.

Block Codes

We now formalise codes that make repeated use of a noisy channel to communicate a predefined set of S messages.
Each $s \in \{1, 2, \ldots, S\}$ is paired with a unique *block* of symbols $\mathbf{x} \in \mathcal{X}^N$.

(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

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

where each $\mathbf{x}^{(s)} \in \mathcal{X}^N$ consists of N symbols from $\mathcal{X}$. The **rate** of such a block code is K/N bits per channel use.

Examples (for Binary Symmetric Channel Q)

- A $(1, 1)$ block code: $\mathcal{S} = \{0, 1\}$ — Rate: 1
- A $(3, 2)$ block code: $\mathcal{S} = \{000, 001, 100, 111\}$ — Rate: $\frac{2}{3}$
- A $(3, \log_2 3)$ block code: $\mathcal{S} = \{001, 010, 100\}$ — Rate: $\frac{\log_2 3}{3} \approx 0.53$

Decoding Block Codes

An (N, K) block code sends each message $s \in \{1, 2, \ldots, 2^K\}$ over a channel Q as $\mathbf{x}^s \in \mathcal{X}^N$ and the block $\mathbf{y} \in \mathcal{Y}^N$ is received. How does the receiver determine which s was transmitted?

Decoding Block Codes

An (N, K) block code sends each message $s \in \{1, 2, \ldots, 2^K\}$ over a channel Q as $\mathbf{x}^s \in \mathcal{X}^N$ and the block $\mathbf{y} \in \mathcal{Y}^N$ is received. How does the receiver determine which s was transmitted?

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

An (N, K) block code sends each message $s \in \{1, 2, \ldots, 2^K\}$ over a channel Q as $\mathbf{x}^s \in \mathcal{X}^N$ and the block $\mathbf{y} \in \mathcal{Y}^N$ is received. How does the receiver determine which s was transmitted?

Block Decoder

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

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

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

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, $d_{opt}(\mathbf{y}) = \arg\max_s P(s|\mathbf{y})$.

Reliability

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

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

Reliability

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

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

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

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

Reliability

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

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

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 $p_B \to 0$.

Achievable Rates

If it is possible to construct codes with rate R for a channel that can have arbitrarily small error the rate R is said to be *achievable*. Formally:

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

If it is possible to construct codes with rate R for a channel that can have arbitrarily small error the rate R is said to be *achievable*. Formally:

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

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

The main "trick" to minimising p_{BM} is to construct a (N, K) block code with (almost) **non-confusable** codes. That is, a code such that the set of $\mathbf{y}$ that each $\mathbf{x}^{(s)}$ are sent to by Q have low probability intersection.

Achievable Rates

If it is possible to construct codes with rate R for a channel that can have arbitrarily small error the rate R is said to be *achievable*. Formally:

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

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

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

The Noisy-Channel Coding Theorem

Informal Statement

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

Informal Statement

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.

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 $1/3 > 0.25$.

The Noisy-Channel Coding Theorem

Informal Statement

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.

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 $1/2 > 0.39 = C$.

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!