Channel Capacity, Sampling Theory and Image, Video & Audio Compression exercises

all numbered exercises are by Cover and Thomas except where noted otherwise

March 11, 2003

Exercise 8.2:
Maximum likelihood decoding. A source produces independent, equally probable symbols from an alphabet $\{a, b\}$ at a rate of one symbol every 3 seconds. These symbols are transmitted over a binary symmetric channel which is used once each second by encoding the source symbol a as 000 and the source symbol b as 111. If, in the corresponding 3 second interval of the channel output, any of the sequences 000, 001, 010, 100 is received then the output is decoded as a; otherwise the output is decoded as b. Let $\epsilon < \frac{1}{2}$ be the channel crossover probability.

(a) For each possible received 3-bit sequence in the interval corresponding to a given source letter, find the probability that a came out of the source given that received sequence.

(b) Using part (a), show that the above decoding rule minimizes the probability of an incorrect decision.

(c) Find the probability of an incorrect decision.

(d) The source is slowed down to produce one letter every $2n + 1$ seconds, a being encoded by $2n + 1$ "0"s and b being encoded by $2n + 1$ "1"s. What decision rule minimizes the probability of error at the decoder? What is the probability of error as $n \to \infty$? What is the transmission rate as $n \to \infty$?

Solution:

(a) We know that the following are true:

$$\begin{aligned}
p(y = 000|x = 000) &= (1 - \epsilon)^3 \\
p(y = 100|x = 000) = p(y = 010|x = 000) = p(y = 001|x = 000) &= (1 - \epsilon)^2\epsilon \\
p(y = 110|x = 000) = p(y = 101|x = 000) = p(y = 011|x = 000) &= (1 - \epsilon)\epsilon^2 \\
p(y = 111|x = 000) &= \epsilon^3
\end{aligned}$$

but we are asked for the conditional probabilities in the other direction, i.e. $p(x|y)$ rather than $p(y|x)$:

$$p(x = 000|y = 000) = \frac{p(x = 000) \times p(y = 000|x = 000)}{p(y = 000)}$$

Now:

$$\begin{aligned}
p(y = 000) &= \sum_x p(x) \times p(y = 000|x) \\
&= p(x = 000) \times p(y = 000|x = 000) + p(x = 111) \times p(y = 000|x = 111) \\
&= \tfrac{1}{2}(1 - \epsilon)^3 + \tfrac{1}{2}\epsilon^3
\end{aligned}$$

So:

$$p(x=000|y=000) = \frac{\frac{1}{2}\times(1-\epsilon)^3}{\frac{1}{2}(1-\epsilon)^3+\frac{1}{2}\epsilon^3} = \frac{(1-\epsilon)^3}{(1-\epsilon)^3+\epsilon^3}$$

Likewise:

$$\begin{aligned} p(x=000|y=001) &= \frac{(1-\epsilon)^2\epsilon}{(1-\epsilon)^2\epsilon+(1-\epsilon)\epsilon^2} \\ &= (1-\epsilon) \\ p(x=000|y=011) &= \frac{(1-\epsilon)\epsilon^2}{(1-\epsilon)^2\epsilon+(1-\epsilon)\epsilon^2} \\ &= \epsilon \\ p(x=000|y=111) &= \frac{\epsilon^3}{(1-\epsilon)^3+\epsilon^3} \end{aligned}$$

$p(x=000|y=001) = p(x=000|y=010) = p(x=000|y=100)$ and
$p(x=000|y=011) = p(x=000|y=101) = p(x=000|y=110)$.

(b) The system is symmetric. Four of the possible cases must be allocated to a, four to case b. 000, 001, 010, and 100 have the highest four probabilities (remember $\epsilon < \frac{1}{2}$) so these are the four which should be allocated to a to minimize the probablity of an incorrect decision.

(c) The probability of an incorrect decision is: $P_e = p(x=a)\times p(y=b|x=a) + p(x=b)\times p(y=a|x=b)$. The symbols are equiprobable ($p(x=a)=p(x=b)=\frac{1}{2}$), $p(y=b|x=a) = p(y=011|x=000)+p(y=101|x=000)+p(y=110|x=000)+p(y=111|x=000)$, and similarly for $p(y=a|x=b)$ so $P_e = 3(1-\epsilon)\epsilon^2+\epsilon^3$.

(d) The decision rule is: if there are n or fewer "1"s, then it is decoded as a, if there are $n+1$ or more "1"s, then it is decoded as b. The probability of error goes to zero as $n\to\infty$. The transmission rate also goes to zero as $n\to\infty$ because we need infinitely many bits to represent a single symbol.

Exercise 8.3:
An additive noise channel. A channel has additive noise, Z, such that the output Y depends on the input X and the noise Z by the rule $Y = X+Z$. The alphabet for x is $\{0,1\}$. The alphabet for Z is $\{a,b\}$ where $\Pr\{Z=a\} = \Pr\{Z=b\} = \frac{1}{2}$, $a,b,\in \mathbb{Z}$. Assume that Z is independent of X. What is the channel capacity of this discrete memoryless channel? [Hint: the channel capacity depends on the value of $b-a$.]
Solution:
$C(X,Y) = \max_{p(X)} I(X;Y)$. If $b-a\in\{-1,1\}$ then Z reduces the channel capacity below one bit, otherwise it has no effect on the transmission as the four possible output symbols (a,b,$a+1$,$b+1$) are all distinct and can therefore be correctly decoded, and therefore the capacity is one bit. In the case $b-a\in\{-1,1\}$, there are three possible output symbols, with probabilities $\left\{\frac{1}{4},\frac{1}{2},\frac{1}{4}\right\}$. For example, consider $a=0$, $b=1$:

$p(y\|x)$	y		
x	0	1	2
0	$\frac{1}{2}$	$\frac{1}{2}$	0
1	0	$\frac{1}{2}$	$\frac{1}{2}$

Here $H(X)=1$, $H(Y)=1.5$, $H(X,Y)=2$, so $C(X,Y)=\max I(X;Y) = H(X)+H(Y)-H(X,Y) = 0.5$.

Our solution is thus: if $b-a\in\{-1,1\}$ then $C(X,Y)=0.5$ otherwise $C(X,Y)=1$.

Exercise 8.9:
The Z channel. The Z channel has binary input and output alphabets $x, y \in \{0, 1\}$ and transition probabilities $p(x|y)$ given by:

	$x = 0$	$x = 1$
$y = 0$	1	0
$y = 1$	$\frac{1}{2}$	$\frac{1}{2}$

Find the input probability distribution which maximises the mutual information and hence determine the channel capacity.
Solution:
Let us say $p = p(x = 1)$. We then either know, or can derive, all of the following facts:
$C(X, Y) = \max_{p(X)} I(X; Y)$
$p(x = 1) = p$ (this is our definition of p)
$p(x = 0) = 1 - p$
$p(y = 1) = 2p$
$p(y = 0) = 1 - 2p$
$p(x = 0, y = 0) = 1 - 2p$
$p(x = 1, y = 0) = 0$
$p(x = 0, y = 1) = p$
$p(x = 1, y = 1) = p$

$$\begin{aligned}
I(X; Y) &= \sum_x \sum_y p(x, y) \log_2 \frac{p(x, y)}{p(x)p(y)} \\
&= (1 - 2p) \log_2 \frac{1 - 2p}{(1 - p)(1 - 2p)} + p \log_2 \frac{p}{2p(1 - p)} + p \log_2 \frac{p}{2p^2} \\
&= (1 - 2p) \log_2 \frac{1}{1 - p} + p \log_2 \frac{1}{2(1 - p)} + p \log_2 \frac{1}{2p} \\
&= (1 - p) \log_2 \frac{1}{1 - p} + p \log_2 \frac{1}{p} - 2p
\end{aligned}$$

By graphing this function, we see that appears to have a maximum at or very near to $p = 0.2$ (I used an Excel spreadsheet to calculate and plot the graph). To find, mathematically, the location of the maximum of this function we need to find where its derivative is equal to zero. First make life easier for ourselves by inverting the contents of the two log functions (this makes the derivative much easier to derive):

$$\begin{aligned}
I(X; Y) &= (1 - p) \log_2 \frac{1}{1 - p} + p \log_2 \frac{1}{p} - 2p \\
&= -(1 - p) \log_2(1 - p) - p \log_2(p) - 2p
\end{aligned}$$

Now take the derivative:

$$\begin{aligned}
\tfrac{dI}{dp} &= \log_2(1 - p) + \tfrac{1}{\log_e 2} - \log_2(p) - \tfrac{1}{\log_e 2} - 2 \\
&= \log_2(1 - p) - \log_2(p) - 2
\end{aligned}$$

and set it equal to zero:

$$\begin{aligned}
& \log_2(1 - p) - \log_2(p) - 2 = 0 \\
\Rightarrow\ & \log_2 \left(\frac{1 - p}{p} \right) = 2 \\
\Rightarrow\ & \frac{1 - p}{p} = 2^2 \\
\Rightarrow\ & 1 - p = 4p \\
\Rightarrow\ & 1 = 5p \\
\Rightarrow\ & p = \tfrac{1}{5}
\end{aligned}$$

so the information content is maximised at $p = 0.2$: $I(X, Y) = 0.322 = C(X, Y)$.

Exercise 8.11:
Zero-error capacity. A channel with alphabet $\{0, 1, 2, 3, 4\}$ has transition probabilities of the form:

$$p(y|x) = \begin{cases} \frac{1}{2}, & y = (x \pm 1)\text{mod}5 \\ 0, & \text{otherwise} \end{cases}$$

That is: any symbol is equally likely to transition to the symbol before it or the symbol after it. For example "2" could be received as "1" or "3" with equal probability.

(a) Compute the theoretical channel capacity in bits.

(b) The zero-error capacity of a channel is the number of bits per channel use that can be transmitted with zero probability of error. Clearly, the zero-erro capacity of this five-symbol channel is at least one bit (transmit "0" or "1" with probability $\frac{1}{2}$). Find a block code that shows that the zero-error capacity is greater than 1 bit. [Hint: consider codes of length 2.]

(c) Estimate the value of the zero-error capacity as the number of symbols in a block goes to infinity. [This is not something that you are likely to be able to calculate, so give your best guess along with some sort of justification.]

Solution:

(a) The problem is nicely symmetric, so we can expect that $I(X; Y)$ will be maximised when all inputs are equiprobable: $p(x = a) = \frac{1}{5}$, $a \in \{0, 1, 2, 3, 4\}$; $p(y = b) = \frac{1}{5}$, $b \in \{0, 1, 2, 3, 4\}$;

$$p(x = a, y = b) = \begin{cases} \frac{1}{10}, & y = (x \pm 1)\text{mod}5 \\ 0, & \text{otherwise} \end{cases}$$

Thus we can compute:

$$\begin{aligned} I(X; Y) &= \sum_x \sum_y p(x, y) \log_2 \frac{p(x, y)}{p(x)p(y)} \\ &= 10 \times \frac{1}{10} \log_2 \frac{1/10}{(1/5)(1/5)} \\ &= \log_2 \frac{25}{10} \\ &= 1.322 \text{ bits} \end{aligned}$$

(b) With one-symbol blocks we can transmit two mutually exclusive codes:

y	0	1	2	3	4
x	1	0	1		1

It is perhaps easier to see what is going on if we reorder the y values to make adjacent values the ones that can be confused:

y	0	2	4	1	3
x	1	1		0	0

For example, $y = 0$ and $y = 2$ are adjacent, both can be generated by sending $x = 1$.

The hint says to try using two-symbol blocks. If we want to transmit one bit per symbol then we need to find four mutually exclusive codes. It should be clear that we could just use 00, 01, 10, 11 to do this. But we want more than one bit per symbol so let's therefore look for some way of sending *five* mutually exclusive codes. We will draw a matrix of the codes which we can receive with every code adjacent to those which can be decoded the same as it:

00	02	04	01	03
20	22	24	21	23
40	42	44	41	43
10	12	14	11	13
30	32	34	31	33

Like Karnaugh maps in IA Digital Electronics, this map wraps around from one side to the other and from top to bottom. We now just need to find non-overlapping squares of four elements. This is the "four mutually exclusive codes" solution corresponding to sending codes 00, 01, 10, 11:

00	02	04	01	*03*
20	22	24	21	*23*
40	42	44	41	*43*
10	12	14	11	*13*
30	*32*	*34*	*31*	*33*

and this is a solution which allows you to send five codes:

00	02	*04*	01	03
20	22	24	21	*23*
40	*42*	44	41	43
10	12	14	*11*	13
30	32	34	31	33

The italicised codes are never used. The five codes which can be sent are: 11, 42, 30, 04, 23.

(c) Obviously, the capacity is going to be at least $\frac{\log_2(5)}{2} = 1.161$ because this is what we found in part (b). Obviously it can be no more than the theoretical channel capacity from part (a). So $1.161 \leq C \leq 1.322$. In the n dimensional case we need to find non-overlapping blocks of size $2 \times 2 \times \cdots \times 2$ in a volume of size $5 \times 5 \times \cdots \times 5$. There will always be *at least* one code which cannot be used because 2^n does not evenly divide into 5^n. It appears that we will probably need to have some proportion of the codes unused in order to fit the small blocks in. I played with the $n = 3$ case for a while but could find no way to include more than 10 non-overlapping blocks. The correct solution to this problem is found by using graph theory, which we aren't actually trained to do. What, therefore, can we say? Is the $n = 2$ case the best that we can do? Does the channel capacity monotonically increase as n increases? Does it bounce up and down as n increases? Is there some magic value of n which gives the best answer? Lovász has written some lecture notes on this (`http://research.microsoft.com/users/lovasz/semidef.ps`, 1MB Postscript file). He states (page 3) that, for any channel, channel capacity does not decrease as n increases (i.e. if I increment n I get either the same or a higher channel capacity). The case under consideration in this exercise (C_5 in Lovász notation) is the simplest case whose limiting channel capacity is not amenable to elementary calculation. Some more complicated calculations (pp. 3–4) show that the channel capacity for this case, as $n \to \infty$ is in fact that which can be achieved for two-symbol blocks. That is: $C = \frac{\log_2(5)}{2}$.

Exercise A [Kuhn]:
JPEG. Which steps of the JPEG (DCT baseline) algorithm cause a loss of information? Distinguish between accidental loss due to rounding errors and information that is removed for a purpose.
Solution:
Accidental loss due to rounding errors from the $RGB \rightarrow YCrCb$ conversion and from the discrete cosine transform. Intentional loss is from the resolution reduction in Cr and Cb planes and from quantisation of DCT results.

Exercise B [Kuhn]:
JPEG. How can you rotate/mirror an already compressed JPEG image without loosing any further information. Why might the resulting JPEG file not have the exact same filelength?
Solution:
Decompress the JPEG file only up to the stage of quantized DCT coefficients. Rotate DCT blocks by negating integer values and swapping coefficients. The reapply zigzag scan, RLE, DC-coefficient DPCM and Huffman encoding. The results of these lossless steps will differ, which can affect the filelength.

Exercise C [Kuhn]:
Fax encoding. Decompress this G3-fax encoded pixel sequence, which starts with a white-pixel count: 1101001011110111101100001101110011010 [Hint: see page 2 of
`http://www.cl.cam.ac.uk/Teaching/2002/InfoTheory/mgk/additional-slides-4up.pdf`
for the decoding table]
Solution:
14 white = 110100
3 black = 10
7 white = 1111
4 black = 011
127 white = 11011 0000110111 00110100

Exercise D [Dodgson]:
Audio & Video encoding. What is it about audio and video data which allows us to use lossy compression for compressing it?
Solution:
The basic answer is that this data is specifically for presentation to the human perceptual systems. These systems have limitations: limited sampling and quantisation resolutions, limited bandwidth, upper and lower limits on how bright or how loud the signal can be. These limitations allow us to intelligently throw away data which the human perceptual systems do not need (or need to a lesser extent than the retained data).

Exercise E [Dodgson]:
Sampling theory. Without using Fourier transforms, show that, for every $\nu_1 > 0$, there exists a ν_2, $0 \leq \nu_2 \leq \nu_b$, such that a sine wave, $y = \sin(2\pi\nu_1 x)$ of frequency ν_1 will produce exactly the same sample values as a sine wave of frequency ν_2 if sampled at the points $\left\{x = \frac{n}{2\nu_b}, n \in \mathbb{Z}\right\}$. Determine a formula for ν_2 in terms of ν_1 and ν_b.
Solution:
We know:

$$\sin(y) = \sin(y + 2\pi k), k \in \mathbb{Z}$$

So we can say:

$$\sin\left(2\pi\nu_1 \frac{n}{2\nu_b}\right) = \sin\left(2\pi n \left(\frac{\nu_1 + 2\nu_b k}{2\nu_b}\right)\right)$$

Implying that the samples will be identical for all $\nu_2 = \nu_1 + 2\nu_b k$, $k \in \mathbb{Z}$. Choose k such that $-\nu_b \leq \nu_2 \leq \nu_b$.

Note that this does *not* always give us a frequency between 0 and ν_b. We can fix this in the following way. We know two more facts:

$$\begin{aligned} \sin(-y) &= -\sin(y) \\ \sin(y+\pi) &= -\sin(y) \end{aligned}$$

Therefore:

$$\sin(y+\pi) = \sin(-y)$$

So, if $-\nu_b \leq \nu_2 < 0$ then replace ν_2 with $\nu_3 = -\nu_2 + \pi = -(\nu_1 + 2\nu_b k) + \pi$. For any frequency ν_1 we are therefore always able to find a frequency between 0 and ν_b which has identical sample values taken at the points $\left\{x = \frac{n}{2\nu_b}, n \in \mathbb{Z}\}\right\}$.