LINKÖPING UNIVERSITY

Solutions to Written Exam in **Image and Audio Coding** **TSBK02**

28th May 2018

1 a) See the course literature.

b) See the course literature.

c) See the course literature.

2 See the course literature.

3 a) A Huffman code for single symbols will give the rate 1.3 bits/symbol and is therefore not enough. We need to code at least two symbols at a time.

One Huffman code (there are others) for pairs is given by

symbols	codeword	codeword length
aa	0	1
ab	100	3
ac	1100	4
ba	101	3
bb	1110	4
bc	11110	5
ca	1101	4
cb	111110	6
cc	111111	6

The code has a mean codeword length of 2.33 bits/codeword and a rate of 1.165 bits/symbol.

b) The intervall corresponding to the sequence is $[0.441,\ 0.4753)$ with size 0.0343. We will need at least $\lceil -\log_2 0.0343 \rceil = 5$ bits in the codeword, maybe one more. Write the two limits as binary numbers:

$$\begin{aligned} 0.441 &= 0.0111000011\ldots \\ 0.4753 &= 0.0111100110\ldots \end{aligned}$$

The smallest five bit binary number inside the interval is 0.01111. However, there are numbers starting with these bits that are outside the interval. Thus, we will need to use six bits.

The codeword will be **011101**.

4 For symmetry reasons, the decision borders must placed as

$$b_0 = -1, \quad b_1 = 0, \quad b_2 = 1$$

The reconstruction point y_2 is given by

$$y_2 = \frac{\int_0^1 x \cdot f_X(x)\ dx}{\int_0^1 \cdot f_X(x)\ dx} = \frac{1/2 - 1/\pi}{1/2} = 1 - \frac{2}{\pi} \approx 0.3634$$

Also for symmetry reasons, $y_1 = -y_2$.

The distortion is given by

$$\begin{aligned} D &= \int_{-1}^{0} (x - y_1)^2 \cdot f_X(x)\ dx + \int_0^1 (x - y_2)^2 \cdot f_X(x)\ dx = \\ &= \int_{-1}^{0} (x + y_2)^2 \cdot f_X(x)\ dx + \int_0^1 (x - y_2)^2 \cdot f_X(x)\ dx = \\ &= \int_{-1}^{1} x^2 \cdot f_X(x)\ dx + y_2^2 \int_{-1}^{1} f_X(x)\ dx - 4y_2 \int_0^1 x \cdot f_X(x)\ dx = \\ &= \frac{4\pi - 12}{\pi^2} \approx 0.05738 \end{aligned}$$

5 We assume that the quantization is fine enough so that we can do the calculations as if the predictor is using the original signal, ie we can disregard the effect of the quantization on the prediction. The variance of the prediction error:

$$\sigma_d^2 = E\{(X_{i,j} - p_{i,j})^2\} \approx E\{(X_{i,j} - a_1 X_{i-1,j} - a_2 X_{i,j-1})^2\}$$

a_1 and a_2 that minimize σ_d^2 are given by

$$\begin{pmatrix} 2209 & 1976 \\ 1976 & 2209 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 2002 \\ 2054 \end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \approx \begin{pmatrix} 0.3730 \\ 0.5962 \end{pmatrix}$$

$$\Rightarrow \sigma_d^2 \approx 237.71$$

Uniform quantization followed by entropy coding to the rate R gives the approximate distortion

$$D \approx \frac{\pi e}{6} \cdot \sigma_d^2 \cdot 2^{-2 \cdot R}$$

or alternatively, with rate as a function of the distortion

$$R \approx \frac{1}{2} \log_2 \frac{\pi e \sigma_d^2}{6D}$$

The signal-to-noise ratio is given by

$$\text{SNR} = 10 \cdot \log_{10} \frac{\sigma_X^2}{D}$$

If we want an SNR of at least 42 dB, we must choose R so that

$$D \leq \frac{\sigma_X^2}{10^{4.2}} \approx 0.1394$$

which gives us the smallest possible rate as approximately 5.62 bits/pixel.

If we didn't use the predictor, we would have needed to use the rate

$$R \approx \frac{1}{2} \log_2 \frac{\pi e \sigma_X^2}{6D} \approx 7.23$$

to reach 42 dB.

6 First we should confirm that the given transform is ON. It is easy to see that all the basis vectors (rows of $\mathbf{A}$) are mutually orthogonal and have length 1 ($2a^2 + 2b^2 = 1$).

Since X_n is gaussian, the transform components will also be gaussian. We need to calculate the variances of the transform components in order to do bit allocation.

$$\begin{aligned}
\sigma_0^2 &= E\{\theta_0^2\} = E\{(aX_0 + bX_1 + bX_2 + aX_3)^2\} = \\
&= R_{XX}(0) + (4ab + 2b^2)R_{XX}(1) + 4abR_{XX}(2) + 2a^2R_{XX}(3) \approx \\
&\approx 0.03980 \\
\sigma_1^2 &= E\{\theta_1^2\} = E\{(bX_0 + aX_1 - aX_2 - bX_3)^2\} = \\
&= R_{XX}(0) + (4ab - 2a^2)R_{XX}(1) - 4abR_{XX}(2) - 2b^2R_{XX}(3) \approx \\
&\approx 0.2297 \\
\sigma_2^2 &= E\{\theta_2^2\} = E\{(bX_0 - aX_1 - aX_2 + bX_3)^2\} = \\
&= R_{XX}(0) - (4ab - 2a^2)R_{XX}(1) - 4abR_{XX}(2) + 2b^2R_{XX}(3) \approx \\
&\approx 0.1891 \\
\sigma_3^2 &= E\{\theta_3^2\} = E\{(aX_0 - bX_1 + bX_2 - aX_3)^2\} = \\
&= R_{XX}(0) - (4ab + 2b^2)R_{XX}(1) + 4abR_{XX}(2) - 2a^2R_{XX}(3) \approx \\
&\approx 3.5408
\end{aligned}$$

Note that since we have a high frequency signal, most of the signal energy is concentrated in the highest frequency transform component.

Check that the average value of the variances is equal to the signal variance.

Alternatively we can find the variances as the diagonal elements of $\mathbf{R}_\theta = \mathbf{A} \cdot \mathbf{R}_X \cdot \mathbf{A}^T$, where

$$\mathbf{R}_X = \begin{pmatrix} 1 & -0.94 & (-0.94)^2 & (-0.94)^3 \\ -0.94 & 1 & -0.94 & (-0.94)^2 \\ (-0.94)^2 & -0.94 & 1 & -0.94 \\ (-0.94)^3 & (-0.94)^2 & -0.94 & 1 \end{pmatrix}$$

The bit allocation that minimizes the distortion is given by $R_0 = 0, R_1 = 2, R_2 = 2, R_3 = 4$ which gives the average distortion

$$D \approx (\sigma_0^2 + 0.1175 \cdot \sigma_1^2 + 0.1175 \cdot \sigma_2^2 + 0.009497 \cdot \sigma_3^2)/4 \approx 0.03067$$

The resulting signal-to-noise ratio is

$$\text{SNR} = 10 \cdot \log_{10} \frac{1}{D} \approx 15.13 \text{ [dB]}$$

Without the transform, we would have gotten the distortion $D \approx 0.1175 \cdot \sigma_X^2 = 0.1175$ and the signal-to-noise ratio

$$\text{SNR} = 10 \cdot \log_{10} \frac{1}{D} \approx 9.30 \text{ [dB]}$$

The gain from the transform is thus 5.83 dB.