LINKÖPING UNIVERSITY

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

25th October 2017

1 a) See the course literature.

b) See the course literature.

c) See the course literature.

d) See the course literature.

e) See the course literature.

2 a) A Huffman code for single symbols will give the rate 1.2 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 1.8675 bits/codeword and a rate of 0.93375 bits/symbol.

b) The intervall corresponding to the sequence is $[0.76,\ 0.7856)$ with size 0.0256. We will need at least $\lceil -\log_2 0.0256 \rceil = 6$ bits in the codeword, maybe one more. The smallest six bit binary number inside the interval is 0.110001, and all numbers starting with these bits are also inside the interval. Thus, six bits are enough and the codeword is thus **110001**.

3 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

$$y_2 = \frac{\int_0^1 x \cdot f_X(x)dx}{\int_0^1 \cdot f_X(x)dx} = \frac{1/8}{1/2} = \frac{1}{4}$$

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

The distortion is given by

$$D = \int_{-1}^{0} (x + \frac{1}{4})^2 \cdot f_X(x)dx + \int_0^1 (x - \frac{1}{4})^2 \cdot f_X(x)dx = \frac{3}{80}$$

4 The quantization is so fine that we can do the approximation that the prediction is made from original signal values and not reconstructed values.

Predictor for the left channel

$$p_i = a \cdot \hat{V}_{i-1} \approx a_1 \cdot V_{i-1}$$

Prediction error variance

$$\sigma_{Ve}^2 = E\{(V_i - p_i)^2\} \approx E\{(V_i - a \cdot V_{i-1})^2\} = (1 + a^2)E\{V_i^2\} - 2a \cdot E\{V_i \cdot V_{i-1}\}$$

Differentiate with respect to a and set equal to 0, which gives us the solution

$$a = \frac{E\{V_i \cdot V_{i-1}\}}{E\{V_i^2\}} \approx 0.8471$$

$$\sigma_{Ve}^2 \approx 0.04802$$

The distorsion is approximately

$$D_V \approx c \cdot \sigma_{Ve}^2 \cdot 2^{-2 \cdot 8}$$

where c depends on the distribution and the choice of quantizer. Approximating the prediction error as being gaussian and that the arithmetic coder gives a rate equal to the memoryless entropy of the quantized prediction error, we will get $c = \frac{\pi e}{6}$ and thus

$$D \approx 1.0430 \cdot 10^{-6}$$

and SNR

$$\text{SNR} = 10 \cdot \log_{10} \frac{\sigma_X^2}{D} \approx 52.12 \text{ dB}$$

Predictor for the right channel

$$q_i = b \cdot \hat{V}_i \approx b \cdot V_i$$

Prediction error variance

$$\sigma_{He}^2 = E\{(H_i - q_i)^2\} \approx E\{(H_i - b \cdot V_i)^2\} = (1+b^2)E\{V_i^2\} - 2b \cdot E\{V_i \cdot H_i\}$$

Differentiate with respect to b and set equal to 0, which gives us the solution

$$b = \frac{E\{V_i \cdot H_i\}}{E\{V_i^2\}} \approx 0.9706$$

$$\sigma_{He}^2 \approx 0.009853$$

Again approximating the prediction error as gaussian, we get the distorsion

$$D_H \approx \frac{\pi e}{6} \cdot \sigma_{He}^2 \cdot 2^{-2 \cdot R_H}$$

Set $D_V = D_H$ and solve for R_H:

$$R_H = 8 - \frac{1}{2} \log_2 \frac{\sigma_{Ve}^2}{\sigma_{He}^2} \approx 6.86$$

Thus, using the rates 8 bits/sample for the left channel and 6.86 bits/sample for for the right channel will give approximately the same distorsion and SNR for both channels.

5 Variances for the four transform components:

$$\begin{aligned}
\sigma_0^2 &= E\{\theta_0^2\} = \frac{1}{4}E\{(X_0 + X_1 + X_2 + X_3)^2\} = \\
&= \frac{1}{4}(4R_{XX}(0) + 6R_{XX}(1) + 4R_{XX}(2) + 2R_{XX}(3)) \approx 3.5699 \\
\sigma_1^2 &= E\{\theta_1^2\} = \frac{1}{20}E\{(3X_0 + X_1 - X_2 - 3X_3)^2\} = \\
&= \frac{1}{20}(20R_{XX}(0) + 10R_{XX}(1) - 12R_{XX}(2) - 18R_{XX}(3)) \approx 0.2799 \\
\sigma_2^2 &= E\{\theta_2^2\} = \frac{1}{4}E\{(X_0 - X_1 - X_2 + X_3)^2\} = \\
&= \frac{1}{4}(4R_{XX}(0) - 2R_{XX}(1) - 4R_{XX}(2) + 2R_{XX}(3)) \approx 0.09369 \\
\sigma_3^2 &= E\{\theta_3^2\} = \frac{1}{20}E\{(X_0 - 3X_1 + 3X_2 - X_3)^2\} = \\
&= \frac{1}{20}(20R_{XX}(0) - 30R_{XX}(1) + 12R_{XX}(2) - 2R_{XX}(3)) \approx 0.05650
\end{aligned}$$

Alternatively you can calculate the variances as the diagonal elements of $\mathbf{A} \cdot \mathbf{R}_X \cdot \mathbf{A}^T$, where

$$\mathbf{R}_X = \begin{pmatrix} 1 & 0.91 & 0.91^2 & 0.91^3 \\ 0.91 & 1 & 0.91 & 0.91^2 \\ 0.91^2 & 0.91 & 1 & 0.91 \\ 0.91^3 & 0.91^2 & 0.91 & 1 \end{pmatrix}$$

The average rate should be 2 bits/sample, so we should allocate $2 \cdot 4 = 8$ total bits to the four transform components. The distortion is minimized if we allocate four bits to θ_0, two bits to θ_1, one bit to θ_2 and one bit to θ_3. The average distortion is

$$D \approx \frac{1}{4}(0.009497 \cdot \sigma_0^2 + 0.1175 \cdot \sigma_1^2 + 0.3634 \cdot \sigma_2^2 + 0.3634 \cdot \sigma_3^2) \approx 0.03034$$

The signal to noise ratio is

$$10 \cdot \log_{10} \frac{\sigma_X^2}{D} = 10 \cdot \log_{10} \frac{1}{D} \approx 15.18 \quad \text{[dB]}$$

If we instead use a Hadamard transform, we have the transform matrix

$$\mathbf{A} = \begin{pmatrix} 1/2 & 1/2 & 1/2 & 1/2 \\ 1/2 & 1/2 & -1/2 & -1/2 \\ 1/2 & -1/2 & -1/2 & 1/2 \\ 1/2 & -1/2 & 1/2 & -1/2 \end{pmatrix}$$

Noting that θ_0 and θ_2 are exactly the same as for the polynomial transform, we ger the transform component variances

$$\begin{aligned}
\sigma_0^2 &\approx 3.5699 \\
\sigma_1^2 &= E\{\theta_1^2\} = \frac{1}{4}E\{(X_0 + X_1 - X_2 - X_3)^2\} = \\
&= \frac{1}{4}(4R_{XX}(0) + 2R_{XX}(1) - 4R_{XX}(2) - 2R_{XX}(3)) \approx 0.2501 \\
\sigma_2^2 &\approx 0.09369 \\
\sigma_3^2 &= E\{\theta_3^2\} = \frac{1}{4}E\{(X_0 - X_1 + X_2 - X_3)^2\} = \\
&= \frac{1}{4}(4R_{XX}(0) - 6R_{XX}(1) + 4R_{XX}(2) - 2R_{XX}(3)) \approx 0.08631
\end{aligned}$$

The optimal bit allocation is the same as for the polynomial transform, which gives the distortion

$$D \approx \frac{1}{4}(0.009497 \cdot \sigma_0^2 + 0.1175 \cdot \sigma_1^2 + 0.3634 \cdot \sigma_2^2 + 0.3634 \cdot \sigma_3^2) \approx 0.03218$$

and the signal to noise ratio

$$10 \cdot \log_{10} \frac{\sigma_X^2}{D} = 10 \cdot \log_{10} \frac{1}{D} \approx 14.92 \quad [\text{dB}]$$

That is, the Hadamard transform is 0.25 dB worse.