LINKÖPING UNIVERSITY

Solutions to Written Exam in **Image and Audio Compression** **TSBK38**

21st March 2024

1 See the course literature.

2 See the course literature.

3
a) See the course literature.
b) See the course literature.
c) See the course literature.

4
a) See the course literature.
b) See the course literature.
c) See the course literature.

5 See the course literature.

6
a) The lowest possible rate is given by the entropy rate of the source. For this memoryless source it is

$$H(X) \approx 2.0087$$

b) Code 3 is not uniquely decodable and is therefore not usable for lossless coding.

Code 1 gives the rate 2.05 bits/symbol for the given source. Code 2 gives the rate 2.2 bits/symbol for the given source.

Thus, code 1 is the best choice for the given source.

7 See the course literature.

8 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**.

9 The distribution is a triangle distribution, symmetric around the origin. This means that the three reconstruction levels are located symmetrically, ie $y_1 = -y_3$, $y_2 = 0$ and y_3. The decision borders should be halfway between the reconstruction levels, ie $b_0 = -1$, $b_1 = -b_2$, $b_2 = \frac{y_3}{2}$ and $b_3 = 1$. Thus, we only have to find one unknown value.

The reconstruction levels should be in the center of gravity of each area, which gives us the equation.

$$y_3 = \frac{\int_{b_2}^{1} x \cdot f_X(x)dx}{\int_{b_2}^{1} f_X(x)dx} = \frac{1 - 3b_2^2 + 2b_2^3}{3 - 6b_2 + 3b_2^2}$$

Since $y_3 = 2b_2$ we finally arrive at the equation

$$4b_2^3 - 9b_2^2 + 6b_2 - 1 = 4(b_2 - 1)^2(b_2 - \frac{1}{4}) = 0$$

where only the root $b_2 = \frac{1}{4}$ is a reasonable solution.

Thus, the decision borders are -1, $-\frac{1}{4}$, $\frac{1}{4}$ and 1. The reconstruction levels are $-\frac{1}{2}$, 0 and $\frac{1}{2}$.

The distortion is given by

$$D = 2\left(\int_0^{1/4} x^2(1-x)dx + \int_{1/4}^{1} (x - \frac{1}{2})^2(1-x)dx\right) = \frac{5}{192}$$

10 5 bits/pixel can be seen as fine quantization, which means that we can ignore the effect of the quantization on the predictor, i.e., we assume that the prediction is made on original values of Y_n. Since we are free to choose the type of quantization too, we of course do uniform quantization followed by entropy coding. Again we use our fine quantization approximation, and assume that the prediction error will be gaussian, giving us the distortion

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

where σ_d^2 is the variance of the prediction error, and $R = 5$.

We first try to use a one-step predictor $p_n = a_1 \cdot \hat{Y}_{n-1}$ and find a_1 that minimize σ_d^2

$$\begin{aligned} \sigma_d^2 &= E\{(Y_n - p_n)^2\} \approx \\ &\approx E\{(Y_n - a_1 \cdot Y_{n-1})^2\} = \\ &= (1 + a_1^2)R_{YY}(0) - 2a_1 \cdot R_{YY}(1) \end{aligned}$$

Differentiate with respect to a_1 and set equal to 0, which gives us the solution

$$a_1 = \frac{R_{YY}(1)}{R_{YY}(0)} \approx 0.9234$$

$$\sigma_d^2 \approx 2.9235$$

The resulting signal to noise ratio is

$$\text{SNR} \approx 10 \cdot \log_{10} \frac{\sigma_Y^2}{\sigma_d^2 \frac{\pi e}{6} 2^{-2R}} \approx 36.9 \text{ [dB]}$$

which is not good enough.

We instead use the predictor $p_n = a_1 \cdot \hat{Y}_{n-1} + a_2 \cdot \hat{Y}_{n-2}$ and find a_1 and a_2 that minimize σ_d^2

$$\begin{aligned}
\sigma_d^2 &= E\{(Y_n - p_n)^2\} \approx \\
&\approx E\{(Y_n - a_1 \cdot Y_{n-1} - a_2 \cdot Y_{n-2})^2\} = \\
&= (1 + a_1^2 + a_2^2) R_{YY}(0) - 2a_1 \cdot R_{YY}(1) - 2a_2 \cdot R_{YY}(2) + 2a_1 a_2 \cdot R_{YY}(1)
\end{aligned}$$

Differentiate with respect to a_1 and a_2 and set equal to 0, which gives us the solution

$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} R_{YY}(0) & R_{YY}(1) \\ R_{YY}(1) & R_{YY}(0) \end{pmatrix}^{-1} \begin{pmatrix} R_{YY}(1) \\ R_{YY}(2) \end{pmatrix} \approx \begin{pmatrix} 1.3613 \\ -0.4742 \end{pmatrix}$$

$$\sigma_d^2 \approx 2.2660$$

The resulting signal to noise ratio is

$$\text{SNR} \approx 10 \cdot \log_{10} \frac{\sigma_Y^2}{\sigma_d^2 \frac{\pi e}{6} 2^{-2R}} \approx 38.0 \text{ [dB]}$$

which is better than the requested 37.5 dB.

Thus, the shortest predictor that solves the problem is a two-step predictor.

11 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.