Solutions to Written Exam in

Image and Audio Coding

TSBK02

20th August 2018

1 a) See the course literature.

b) See the course literature.

c) See the course literature.

d) See the course literature.

2 a) See the course literature.

b) See the course literature.

3 a) The lowest possible rate is given by the entropy rate of the source. For a memoryless source this is just the entropy of single symbols:

$$\begin{aligned} H(X) &= -0.5 \log 0.5 - 0.4 \log 0.4 - 0.1 \log 0.1 \\ &\approx 1.3610 \end{aligned}$$

b) One Huffman code (there are others) is given by

symbols	codeword	codeword length
aa	00	2
ab	10	2
ac	01100	5
ba	11	2
bb	010	3
bc	01110	5
ca	01101	5
cb	011110	6
cc	011111	6

The code has a mean codeword length of 2.78 bits/codeword and a rate of 1.39 bits/symbol.

4 a) Number of quantization levels

$$M = 2^{12} = 4096$$

The step length of the quantizer is

$$\Delta = \frac{10\sigma}{M}$$

Distortion

$$D \approx \frac{\Delta^2}{12} = \frac{100\sigma^2}{12M^2}$$

Signal-to-noise ratio

$$10 \log_{10} \frac{\sigma^2}{D} \approx 10 \log_{10} 2013265.92 \approx 63.0 \quad [\text{dB}]$$

b) The theoretical limit is given by the rate-distortion function. For a memoryless gaussian signal, it is

$$R(D) = \frac{1}{2} \log_2 \frac{\sigma^2}{D}$$

or equivalently, the distortion-rate function is

$$D(R) = \sigma^2 \cdot 2^{-2R}$$

In our case, $R = 12$. This gives us the SNR

$$10 \log_{10} \frac{\sigma^2}{D} = 10 \log_{10} 16777216 \approx 72.2 \quad [\text{dB}]$$

5 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 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

$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} R_{YY}(0) & R_{YY}(1) \\ R_{YY}(1) & R_{YY}(0) \end{pmatrix}^{-1} \begin{pmatrix} R_{YY}(1) \\ R_{YY}(2) \end{pmatrix} \approx \begin{pmatrix} 1.4993 \\ -0.7987 \end{pmatrix}$$

$$\sigma_d^2 \approx 1.0023$$

The resulting signal to noise ratio is

$$\text{SNR} \approx 10 \cdot \log_{10} \frac{\sigma_Y^2}{\sigma_d^2 \frac{\pi e}{6} 2^{-2R}} \approx 38.1 \text{ [dB]}$$

which is better than the requested 35 dB, thus we have solved the problem.

6 The transform matrix (basis vectors in the rows) of a 3 point DCT is

$$\mathbf{A} = \sqrt{\frac{2}{3}} \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 1/\sqrt{2} \\ \sqrt{3}/2 & 0 & -\sqrt{3}/2 \\ 1/2 & -1 & 1/2 \end{pmatrix} = \begin{pmatrix} 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \\ 1/\sqrt{2} & 0 & -1/\sqrt{2} \\ 1/\sqrt{6} & -2/\sqrt{6} & 1/\sqrt{6} \end{pmatrix}$$

Variances of the three transform components:

$$\begin{aligned}
\sigma_0^2 &= E\{\theta_0^2\} = \frac{1}{3}E\{(L_i + C_i + R_i)^2\} = \\
&= \frac{1}{3}(3 + 2 \cdot 0.93 + 2 \cdot 0.9 + 2 \cdot 0.93) = 2.84 \\
\sigma_1^2 &= E\{\theta_1^2\} = \frac{1}{2}E\{(L_i - R_i)^2\} = \\
&= \frac{1}{2}(2 - 2 \cdot 0.9) = 0.1 \\
\sigma_2^2 &= E\{\theta_2^2\} = \frac{1}{6}E\{(L_i - 2C_i + R_i)^2\} = \\
&= \frac{1}{6}(6 - 4 \cdot 0.93 + 2 \cdot 0.9 - 4 \cdot 0.93) = 0.06
\end{aligned}$$

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

$$\mathbf{R}_X = \begin{pmatrix} E\{L_i^2\} & E\{L_iC_i\} & E\{L_iR_i\} \\ E\{L_iC_i\} & E\{C_i^2\} & E\{R_iC_i\} \\ E\{L_iR_i\} & E\{R_iC_i\} & E\{R_i^2\} \end{pmatrix} = \begin{pmatrix} 1 & 0.93 & 0.9 \\ 0.93 & 1 & 0.93 \\ 0.9 & 0.93 & 1 \end{pmatrix}$$

The desired rate is 2 bits/sample/channel, so we should allocate a total of $2 \cdot 3 = 6$ bits to the three transform components. The distortion is minimized if we give 4 bits to component 0, and one bit each to the other two components. The average distortion is

$$D \approx \frac{1}{3}(0.009497 \cdot \sigma_0^2 + 0.3634 \cdot \sigma_1^2 + 0.3634 \cdot \sigma_2^2) \approx 0.02837$$

Without transform we get the distortion

$$D_Q \approx \frac{1}{3} \cdot 3 \cdot 0.1175 \cdot 1 = 0.1175$$