LINKÖPING
UNIVERSITY

Solutions to Written Exam in
Image and Audio Compression
TSBK38

23rd March 2023

1 Some symbols (or partial sequences of symbols) will be more common than others. If we have a random model for the source this corresponds to having different probabilities. By having short codewords for common symbols and long codewords for uncommon symbols we can get a lower data rate than if we used the same length codewords for all symbols.

Most sources have some kind of memory (dependence between neighbouring symbols in the sequence). This can be used to achieve a lower data rate than if we didn't take the memory into account.

2 See the course literature.

3 See the course literature.

4 a) See the course literature.

b) See the course literature.

c) See the course literature.

5 a) See the course literature.

b) See the course literature.

c) See the course literature.

6 A Huffman code for single symbols will give the rate 1.4 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 2.67 bits/codeword and a rate of 1.335 bits/symbol.

7 See the course literature.

8 If we assume that the x intervall is always closest to 0, the intervall corresponding to the sequence $xyyyy$ is then $[0.68336, 0.8)$. If you ordered your symbols differently, you should at least get an intervall of the same size (0.11664).

We need at least $\lceil -\log 0.11664 \rceil = 4$ bits to specify this intervall. If we write the limits as binary numbers we get

$$\begin{aligned} 0.68336 &= 0.10101110\ldots \\ 0.8 &= 0.11001100\ldots \end{aligned}$$

The smallest binary number with four bits in this interval is 0.1011 We can see that four bits will be enough (all numbers starting with these four bits are also inside the interval). The codeword is thus **1011**.

9 The number of quantization levels in the quantizer is

$$M = \frac{2}{\Delta} = 2 \cdot 2^k = 2^{k+1}$$

The rate after fixed length coding is

$$R = \log M = k + 1 \implies k = R - 1$$

The distortion of the quantizer is

$$D = \frac{\Delta^2}{12} = \frac{2^{-2k}}{12} = \frac{1}{3}2^{-2R}$$

10 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.

11 Transform matrix for a 4 point DWHT:

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

Variances for the four transform components:

$$\begin{aligned}
\sigma_0^2 &= E\{\theta_0^2\} = \frac{1}{4}E\{(X_0+X_1+X_2+X_3)^2\} = \\
&= \frac{1}{4}(4R_{XX}(0)+6R_{XX}(1)+4R_{XX}(2)+2R_{XX}(3)) \approx 3.6157 \\
\sigma_1^2 &= E\{\theta_1^2\} = \frac{1}{4}E\{(X_0+X_1-X_2-X_3)^2\} = \\
&= \frac{1}{4}(4R_{XX}(0)+2R_{XX}(1)-4R_{XX}(2)-2R_{XX}(3)) \approx 0.2243 \\
\sigma_2^2 &= E\{\theta_2^2\} = \frac{1}{4}E\{(X_0-X_1-X_2+X_3)^2\} = \\
&= \frac{1}{4}(4R_{XX}(0)-2R_{XX}(1)-4R_{XX}(2)+2R_{XX}(3)) \approx 0.08294 \\
\sigma_3^2 &= E\{\theta_3^2\} = \frac{1}{4}E\{(X_0-X_1+X_2-X_3)^2\} = \\
&= \frac{1}{4}(4R_{XX}(0)-6R_{XX}(1)+4R_{XX}(2)-2R_{XX}(3)) \approx 0.07706
\end{aligned}$$

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

$$\mathbf{R}_X = \begin{pmatrix} 1 & 0.92 & 0.92^2 & 0.92^3 \\ 0.92 & 1 & 0.92 & 0.92^2 \\ 0.92^2 & 0.92 & 1 & 0.92 \\ 0.92^3 & 0.92^2 & 0.92 & 1 \end{pmatrix}$$

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

$$D \approx \frac{1}{4}(0.009497 \cdot \sigma_0^2 + 0.1175 \cdot \sigma_1^2 + 0.3634 \cdot \sigma_2^2 + 0.3634 \cdot \sigma_3^2) \approx 0.02971$$

The signal to noise ratio is

$$10 \cdot \log_{10} \frac{\sigma_X^2}{D} = 10 \cdot \log_{10} \frac{1}{D} \approx 15.27 \quad [\text{dB}]$$

If we don't use a transform, the distortion is

$$D \approx 0.1175 \cdot \sigma_X^2$$

with signal to noise ratio

$$10 \cdot \log_{10} \frac{\sigma_X^2}{D} = 10 \cdot \log_{10} \frac{1}{D} \approx 9.30 \quad [\mathrm{dB}]$$

Thus, we gain approximatively 5.97 dB by using transform coding.