LINKÖPINGS UNIVERSITET

Lösningsförslag till tentamen i **Kompression av ljud och bild TSBK35**

29:e maj 2023

1 Se kurslitteraturen

2 I en likformig kvantiserare är alla beslutsområden lika stora och rekonstruktionspunkterna ligger mitt i respektive område.

I en Lloyd-Max-kvantiserare lägger man ut beslutsgränser och rekonstruktionspunkter så att distorsionen minimeras.

En Lloyd-Max-kvantiserare ger alltså en lägre distorsion än en likformig kvantiserare med lika många nivåer.

Likformig kvantisering är dock mycket enklare att implementera än Lloyd-Max-kvantisering, och om man gör någon form av variabellängdskodning efter kvantiseringen är det oftast bättre att använda likformig kvantisering.

3 a) Se kurslitteraturen

b) Se kurslitteraturen

4 Se kurslitteraturen

5 Se kurslitteraturen

6 a) Entropin är

$$-0.65 \cdot \log_2 0.65 - 0.25 \cdot \log_2 0.25 - 0.1 \cdot \log_2 0.1 \approx 1.2362 \text{ [bitar]}$$

b) Exempel på huffmankod för par av symboler

par	sannolikhet	längd	kodord
aa	0.4225	1	0
ab	0.1625	3	100
ac	0.065	4	1100
ba	0.1625	3	101
bb	0.0625	4	1110
bc	0.025	5	11110
ca	0.065	4	1101
cb	0.025	6	111110
cc	0.01	6	111111

Kodordsmedelängden är 2.5025 bitar/kodord och medeldatatakten är 1.25125 bitar/symbol. Jämför med entropin från a).

7 För minimal distorsion ska beslutsgränserna läggs mitt emellan rekonstruktionspunkterna och rekonstruktionspunkterna ska ligga i tyngdpunkterna av sina områden. Eftersom fördelningen är symmetrisk och vi bara har två rekonstruktionspunkter får vi beslutsgränserna direkt:

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

Pga symmetrin får vi även att $y_1 = -y_2$. Återstår bara att räkna ut

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

$$\int_0^1 f_X(x)dx = \frac{1}{2}$$

$$\int_0^1 x \cdot f_X(x)dx = \frac{3}{2}\int_0^1 (x^3 - 2x^2 + x)dx = \frac{1}{8}$$

Rekonstruktionspunkterna blir alltså

$$y_1 = -\frac{1}{4}, \quad y_2 = \frac{1}{4}$$

8 Transformmatrisen (basvektorer i raderna) för en 3-punkters DCT ser ut som:

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

Varianser för de tre transformkomponenterna:

$$
\begin{aligned}
\sigma_0^2 &= E\{\theta_0^2\} = \frac{1}{3}E\{(X_0 + X_1 + X_2)^2\} = \\
&= \frac{1}{3}(3R_{XX}(0) + 4R_{XX}(1) + 2R_{XX}(2)) = 5.56 \\
\sigma_1^2 &= E\{\theta_1^2\} = \frac{1}{2}E\{(X_0 - X_2)^2\} = \\
&= \frac{1}{2}(2R_{XX}(0) - 2R_{XX}(2)) = 0.38 \\
\sigma_2^2 &= E\{\theta_2^2\} = \frac{1}{6}E\{(X_0 - 2X_1 + X_2)^2\} = \\
&= \frac{1}{6}(6R_{XX}(0) - 8R_{XX}(1) + 2R_{XX}(2)) = 0.06
\end{aligned}
$$

Alternativt kan man räkna ut varianserna som diagonalelementen i $\mathbf{A} \cdot \mathbf{R}_X \cdot \mathbf{A}^T$, där

$$
\mathbf{R}_X = \begin{pmatrix} 2.00 & 1.86 & 1.62 \\ 1.86 & 2.00 & 1.86 \\ 1.62 & 1.86 & 2.00 \end{pmatrix}
$$

Medeldatatakten ska vara 2 bitar/sampel, så vi ska dela ut $2 \cdot 3 = 6$ bitar på de tre transformkomponenterna. Distorsionen minimeras om man ger 4 bitar till komponent 0, två bitar till komponent 1 och inga bitar till komponent 2. Medeldistorsionen blir

$$
D \approx \frac{1}{3}(0.009497 \cdot \sigma_0^2 + 0.1175 \cdot \sigma_1^2 + \sigma_2^2) \approx 0.0525
$$

Signal-brus-förhållandet blir

$$
10 \cdot \log_{10} \frac{\sigma_X^2}{D} = 10 \cdot \log_{10} \frac{2.00}{D} \approx 15.81 \text{ [dB]}
$$

Utan transform får man distorsionen

$$
D_Q \approx 0.1175 \cdot 2.00 = 0.235
$$

som ger signal-brus-förhållandet

$$
10 \cdot \log_{10} \frac{2.0}{D_Q} \approx 9.30 \text{ [dB]}
$$

9 Prediktorkoefficenterna fås som lösning till ekvationssystemet

$$
\begin{pmatrix} E\{X_{i-1,j}^2\} & E\{X_{i,j-1} \cdot X_{i-1,j}\} \\ E\{X_{i,j-1} \cdot X_{i-1,j}\} & E\{X_{i,j-1}^2\} \end{pmatrix} \cdot \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} E\{X_{i,j} \cdot X_{i-1,j}\} \\ E\{X_{i,j} \cdot X_{i,j-1}\} \end{pmatrix}
$$

$$\begin{pmatrix} 121 & 102 \\ 102 & 121 \end{pmatrix} \cdot \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 109 \\ 112 \end{pmatrix}$$

dvs

$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \approx \begin{pmatrix} 0.4166 \\ 0.5745 \end{pmatrix}$$

Prediktionsfelets varians:

$$\sigma_e^2 = 121 - \begin{pmatrix} a_1 & a_2 \end{pmatrix} \begin{pmatrix} 109 \\ 112 \end{pmatrix} \approx 11.254$$

Givet att kvantiseringen är fin så kommer prediktionsfelet approximativt att vara normalfördelat. Om vi gör likformig kvantisering följt av perfekt minnesfri källkodning blir den resulterande distorsionen

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

och signalbrusförhållandet

$$10 \cdot \log_{10} \frac{\sigma_X^2}{D} = 10 \cdot \log_{10} \frac{121}{D}$$

För att få ett SNR som är minst 36 dB, så måste alltså distorsionen vara högst $121/10^{3.6} \approx 0.03039$. För att detta ska vara uppfyllt måste vi välja R enligt

$$R > \frac{1}{2} \log_2 \frac{\pi e \cdot \sigma_e^2 \cdot 10^{3.6}}{6 \cdot 121} \approx 4.52 \text{ [bitar/pixel]}$$