Predictive Coding

- Lossless predictive coding
- Optimum predictors
- JPEG-LS lossless compression standard
- Lossy predictive coding: DPCM
- Rate distortion performance of DPCM
Lossless Predictive Coding

- Prediction $\mu_p[n]$ is calculated for $x[n]$ from previous samples $x_{N+n}$
- $e[n]$ is prediction error, with greatly reduced statistical dependencies between adjacent samples
- Entropy coder may assume i.i.d. prediction error $e[n]$
- Receiver can reconstruct $x[n]$ without loss for amplitude-discrete signals $x, \mu_p, e \in \mathbb{Z}$
- Much simpler than context-adaptive coder
Lossless Predictive Coding

- Prediction $\mu_p[n]$ is calculated for $x[n]$ from previous samples $x_{N+n}$
- $e[n]$ is prediction error, with greatly reduced statistical dependencies between adjacent samples
- Entropy coder may assume i.i.d. prediction error $e[n]$  
- Receiver can reconstruct $x[n]$ without loss for amplitude-discrete signals $x, \mu_p, e \in \mathbb{Z}$
- Much simpler than context-adaptive coder
Prediction example: test pattern
Prediction example: Cameraman

original

0 0 0
0.95

0

0 0 0

0.95

0

0.5

0

0.5
Histograms: Cameraman

Image signal

Prediction error

-50 0 50

0 0.5 0

x $10^4$

0 0.5 1 1.5 2

0 500 1000 1500 2000 2500 3000
Entropy and variance of the prediction error

- **Approximation of the entropy of the prediction error** $E$

  $$H(E) \approx \log_2 \frac{\sigma_E}{\Delta} + c_{pdf} \quad \text{for} \quad \Delta \ll \sigma_E$$

  - Standard deviation of $E$
  - Constant that depends on the shape of the underlying PDF
  - Quantization step size

- **Shape constant**

  - Gaussian PDF: $c_{pdf} = 2.047$ bit
  - Laplacian PDF: $c_{pdf} = 1.943$ bit

- **With linear prediction of image signals** the prediction error PDF is typically Laplacian.

- **Minimization of prediction error variance or prediction error entropy** typically lead to very similar results.
Optimum linear predictors for the luminance signal $Y$

$$\Delta \ll \sigma_E \Rightarrow H(E) \approx \log_2 \left( \sqrt{2} e \sigma_E \right) - \log_2 \Delta \Rightarrow \arg \min \sigma_E \approx \arg \min H(E)$$

<table>
<thead>
<tr>
<th>$H(X_0)$ [bit]</th>
<th>Predictor</th>
<th>MSE</th>
<th>$H(E)$ [bit]</th>
<th>Criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$a_1$</td>
<td>$a_2$</td>
<td>$a_3$</td>
<td></td>
</tr>
<tr>
<td>7.23</td>
<td>0.595</td>
<td>-0.434</td>
<td>0.831</td>
<td>33.810</td>
</tr>
<tr>
<td></td>
<td>0.464</td>
<td>-0.264</td>
<td>0.799</td>
<td>35.075</td>
</tr>
</tbody>
</table>

Image: ‘Lena’, 512 x 512 pixels, 8 bpp

$\Delta = 1$, $2^8$ levels (-128..127)
Optimum linear predictors for the luminance signal Y

\[ Y \]

\[ x_0 \]

\[ x_1 \]

\[ x_2 \]

\[ x_3 \]

Line of pixels above

Current line of pixels

Constraint: 3 bit word length of the prediction coefficients, +1 bit for sign

<table>
<thead>
<tr>
<th>( H(X_0) ) [bit]</th>
<th>Predictor</th>
<th>MSE</th>
<th>( H(E) ) [bit]</th>
<th>Criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.23</td>
<td>( a_1 ) 5/8</td>
<td>( a_2 ) -1/2</td>
<td>( a_3 ) 7/8</td>
<td>34.161</td>
</tr>
<tr>
<td></td>
<td>( a_1 ) 1/2</td>
<td>( a_2 ) -1/4</td>
<td>( a_3 ) 3/4</td>
<td>35.395</td>
</tr>
</tbody>
</table>

Image: ‘Lena’, 512 x 512 pixels, 8 bpp

\[ \Delta = 1, 2^8 \text{ levels (-128..127)} \]
JPEG-LS lossless compression standard

- Not to be confused with lossless mode of original JPEG
- Based on LOCO-I (Low Complexity Compression of Images) [Weinberger, Seroussi, Sapiro, 1996]
- Predictive coding with nonlinear predictor
- Context-adaptive Golomb coding of prediction error
- 365 different coding contexts, based on pixel differences in the causal neighborhood
- Switches to 1-d run-length coding for one context
- Run-lengths encoded by Golomb code
- “Near-lossless” mode extension
JPEG-LS blockdiagram

[Weinberger, Seroussi, Sapiro, 2000]
JPEG-LS nonlinear predictor

\[ \mu_p = \begin{cases} 
\min(S_1, S_3) & \text{if } S_2 = \max(S_1, S_2, S_3) \\
\max(S_1, S_3) & \text{if } S_2 = \min(S_1, S_2, S_3) \\
S_1 - S_2 + S_3 & \text{else}
\end{cases} \]
JPEG-LS nonlinear predictor

\[ \mu_p = \begin{cases} 
\min(S_1, S_3) & \text{if } S_2 = \max(S_1, S_2, S_3) \\
\max(S_1, S_3) & \text{if } S_2 = \min(S_1, S_2, S_3) \\
S_1 - S_2 + S_3 & \text{else}
\end{cases} \]
JPEG-LS nonlinear predictor

\[
\mu_p = \begin{cases} 
\min(S_1, S_3) & \text{if } S_2 = \max(S_1, S_2, S_3) \\
\max(S_1, S_3) & \text{if } S_2 = \min(S_1, S_2, S_3) \\
S_1 - S_2 + S_3 & \text{else}
\end{cases}
\]
JPEG-LS context labeling

- Quantize each $\Delta_i$ to 1 out of 9 different indices: $9^3 = 729$ distinct contexts, default thresholds 3, 7, 21
- Further reduce to 365 contexts by exploiting sign symmetries
- For $\Delta_1 = \Delta_2 = \Delta_3 = 0$, switch to run-length coding
Lossy Predictive Coding: Differential Pulse Code Modulation (DPCM)

Reconstruction error = quantization error

\[ x' - x = e' - e = q \]
Quantization error feedback in the DPCM coder

- For a linear predictor, the DPCM coder is equivalent to:

\[
\begin{align*}
\text{Image filtered by} & \quad [1 - P(\omega_x, \omega_y)] \\
\text{Quantization error filtered by} & \quad [1 - P(\omega_x, \omega_y)]
\end{align*}
\]

- Linear DPCM decoder

\[
\frac{1}{1 - P(\omega_x, \omega_y)}
\]
Example of intraframe DPCM coding

- Slope overload
- Edge busyness
- Granular noise

- Linear predictor:
  - 0
  - \( \frac{1}{4} \)
  - \( \frac{1}{4} \)
  - \( \frac{1}{2} \)

- Lloyd-Max quantizers
- Fixed-length coding

1 bit/pixel: prediction error coding
2 bit/pixel: edge busyness
3 bit/pixel
4 bit/pixel
Original
Signal distortions due to intraframe DPCM coding

- **Granular noise**: random noise in flat areas of the picture
- **Edge busyness**: jittery appearance of edges (for video)
- **Slope overload**: blur of high-contrast edges, Moire patterns in periodic structures.

[Netravali + Haskell]
DPCM with entropy-constrained quantization

\[ K = 511 \]
\[ H(e') = H(e) = 4.79 \text{ bpp} \]

\[ K = 15 \]
\[ H(e') = 1.98 \text{ bpp} \]

\[ K = 3 \]
\[ H(e') = 0.88 \text{ bpp} \]

\( K \) ... number of reconstruction levels,
\( H(e') \) ... entropy of quantized prediction error

[J. R. Ohm]
Recall from Chapter “Quantization”
High-rate performance of scalar quantizers

- High-rate distortion-rate function

\[ d(R) \approx \varepsilon^2 \sigma_X^2 2^{-2R} \]

- Scaling factor \( \varepsilon^2 \)

<table>
<thead>
<tr>
<th></th>
<th>Shannon LowBd</th>
<th>Lloyd-Max</th>
<th>Entropy-coded</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uniform</td>
<td>( \frac{6}{\pi e} \approx 0.703 )</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Laplacian</td>
<td>( \frac{e}{\pi} \approx 0.865 )</td>
<td>( \frac{9}{2} = 4.5 )</td>
<td>( \frac{e^2}{6} \approx 1.232 )</td>
</tr>
<tr>
<td>Gaussian</td>
<td>1</td>
<td>( \frac{\sqrt{3\pi}}{2} \approx 2.721 )</td>
<td>( \frac{\pi e}{6} \approx 1.423 )</td>
</tr>
</tbody>
</table>
Predictive coding gain

- High-rate distortion-rate function with DPCM

\[ d_{DPCM}(R) \approx \epsilon_E^2 \sigma_E^2 2^{-2R} \]

- Prediction gain

\[ G_{DPCM} = \frac{\epsilon_X^2 \sigma_X^2}{\epsilon_E^2 \sigma_E^2} \]

- Linear prediction: smallest achievable prediction error variance for \( N \)-dimensional signal determined by spectral flatness

\[ \frac{\sigma_E^2}{\sigma_X^2} = \frac{1}{\sigma_X^2} \exp \left( \frac{1}{(2\pi)^N} \int_{\Omega} \ln(\Phi_{XX}(\Omega)) d\Omega \right) \]
Example

- 1-D Gaussian Markov-1 process with correlation coefficient $\rho$
- Autocorrelation function $E[X_n X_{n-k}] = \sigma_X^2 \rho^{|k|}$
- Prediction gain $G_{DPCM} = \frac{1}{1 - \rho^2}$
R-D curves for Gauss-Markov-1 source

\[ \text{SNR [dB]} = 10 \cdot \log_{10} \frac{\sigma_x^2}{D} \]

- Linear predictor, order \( N=1, a=0.9 \)
- Entropy-Constrained Scalar Quantizer with Huffman VLC
- Iterative design algorithm applied

\[ 10 \cdot \log_{10} \left( \frac{\pi e}{6} \right) = 1.53 \text{ dB} \]
\[ 10 \cdot \log_{10} \left( \frac{1}{1 - \rho^2} \right) = 7.2 \text{ dB} \]
• Wiegand, Schwarz, Chapter 6
• Taubman, Marcellin, 2.4.2, 3.3, Chapter 20 (JPEG-LS)