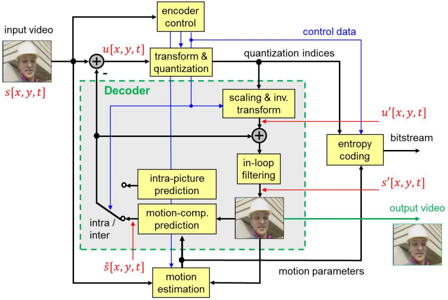
Source Coding and Compression

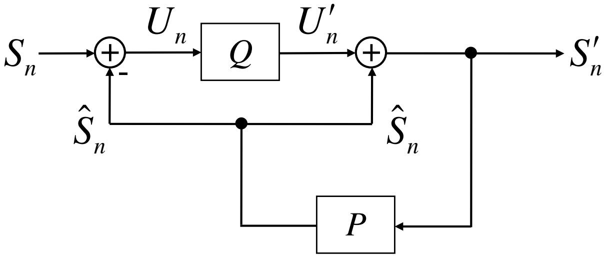
Heiko Schwarz



Contact:

Dr.-Ing. Heiko Schwarz
heiko.schwarz@hhi.fraunhofer.de

Predictive Coding



Outline

Part I: Source Coding Fundamentals

- Probability, Random Variables and Random Processes
- Lossless Source Coding
- Rate-Distortion Theory
- Quantization
- **Predictive Coding**
  - Optimal Prediction
  - Linear and Affine Prediction
  - Predictive Coding: DPCM
- Transform Coding

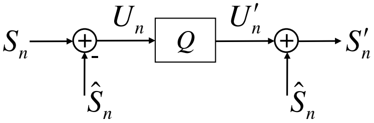
Part II: Application in Image and Video Coding

- Still Image Coding / Intra-Picture Coding
- Hybrid Video Coding (From MPEG-2 Video to H.265/HEVC)

Predictive Coding – Introduction

- The better the future of a random process is predicted from the past and the more redundancy the signal contains, the less new information is contributed by each successive observation of the process

- Predictive coding idea:
  1. Predict a sample using an estimate which is a function of past samples
  2. Quantize residual between signal and its prediction
  3. Add quantizer residual and prediction to obtain reconstructed sample



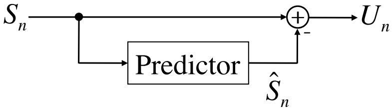
Problems:

- How to obtain the predictor $\hat{S}_n$?
- How to combine predictor and quantizer?

Prediction

- Statistical estimation procedure:

  Value of random variable S_n of random process $\{S_n\}$ is estimated using values of other random variables of the random process



- Select: Set of observed random variables $\mathcal{B}_n$

  $\Longrightarrow$ Typical example: N random variables that directly precede S_n

$$\mathcal{B}_n = \{S_{n-1}, S_{n-2}, \cdots, S_{n-N}\} \quad (441)$$

- Predictor for S_n: Deterministic function of observation set $\mathcal{B}_n$

$$\boxed{\hat{S}_n = A_n(\mathcal{B}_n)} \quad (442)$$

- Prediction error

$$\boxed{U_n = S_n - \hat{S}_n = S_n - A_n(\mathcal{B}_n)} \quad (443)$$

Prediction Performance

MSE distortion using $u_i = s_i - \hat{s}_i$ and $s'_i = u'_i + \hat{s}_i$

$$d_N(\boldsymbol{s}, \boldsymbol{s'}) = \frac{1}{N}\sum_{i=0}^{N-1}(s_i - s'_i)^2 = \frac{1}{N}\sum_{i=0}^{N-1}(u_i + \hat{s}_i - u'_i - \hat{s}_i)^2 = d_N(\boldsymbol{u}, \boldsymbol{u'}) \quad (444)$$

$\Longrightarrow$ Operational rate-distortion function of a predictive coding is equal to op. r-d function of (scalar) quantization of the prediction residuals

Operational distortion-rate function: $D(R) = \sigma_U^2 \cdot g(R)$

- σ_U^2: variance of the prediction residual
- $g(R)$: depends only on the type of the distribution of the residuals

$\Longrightarrow$ Assume stationary processes: $A_n(\cdot)$ becomes $A(\cdot)$

$\Longrightarrow$ Neglect the dependency on the distribution type

$\Longrightarrow$ Define: Predictor $A(\mathcal{B}_n)$ given an observation set $\mathcal{B}_n$ is optimal if it minimizes variance σ_U^2

Optimal Prediction

- Optimization criterion typically used in literature:

$$\boxed{\epsilon_U^2 = E\{U_n^2\} = E\Big\{\big(S_n - \hat{S}_n\big)^2\Big\} = E\Big\{\big(S_n - A(\mathcal{B}_n)\big)^2\Big\}} \tag{445}$$

- Minimization of second moment

$$\begin{aligned} \epsilon_U^2 &= E\{(U_n - \mu_U + \mu_U)^2\} \\ &= E\{(U_n - \mu_U)^2\} + 2E\{(U_n - \mu_U)\mu_U\} + E\{\mu_U^2\} \\ &= \sigma_U^2 + \mu_U^2 + 2\mu_U(E\{U_n\} - \mu_U) \\ &= \sigma_U^2 + \mu_U^2 \end{aligned} \tag{446}$$

implies minimization of variance σ_U^2 and mean μ_U

- Solution: Conditional mean (see proof in [Wiegand and Schwarz])

$$\boxed{\hat{S}_n^* = A^*(\mathcal{B}_n) = E\{S_n \,|\, \mathcal{B}_n\}} \tag{447}$$

$\Longrightarrow$ General case requires storage of large tables

Optimal Prediction for Autoregressive Processes

- Autoregressive process of order m (AR(m) process)

$$\begin{aligned} S_n &= Z_n + \mu_S + \sum_{i=1}^{m} a_i \cdot (S_{n-i} - \mu_S) \\ &= Z_n + \mu_S \cdot (1 - \boldsymbol{a}_m^T \boldsymbol{e}_m) + \boldsymbol{a}_m^T \boldsymbol{S}_{n-1}^{(m)} \end{aligned} \quad (448)$$

  where
  - $\{Z_n\}$ is a zero-mean iid process
  - μ_S is the mean of the AR(m) process
  - $\boldsymbol{a}_m = (a_1, \cdots, a_m)^T$ is a constant parameter vector
  - $\boldsymbol{e}_m = (1, \cdots, 1)^T$ is an m-dimensional unit vector

- Prediction of S_n given the vector $\boldsymbol{S}_{n-1} = (S_{n-1}, \cdots, S_{n-N})$ with $N \geq m$

$$\begin{aligned} E\{S_n \,|\, \boldsymbol{S}_{n-1}\} &= E\big\{Z_n + \mu_S(1 - \boldsymbol{a}_N^T \boldsymbol{e}_N) + \boldsymbol{a}_N^T\, \boldsymbol{S}_{n-1} \,|\, \boldsymbol{S}_{n-1}\big\} \\ &= \mu_S(1 - \boldsymbol{a}_N^T \boldsymbol{e}_N) + \boldsymbol{a}_N^T\, \boldsymbol{S}_{n-1} \end{aligned} \quad (449)$$

  where $\boldsymbol{a}_N = (a_1, \cdots, a_m, 0, \cdots, 0)^T$

Affine Prediction

- Suitable structural constraint: Affine predictor

$$\boxed{\hat{S}_n = A(\boldsymbol{S}_{n-k}) = h_0 + \boldsymbol{h}_N^T \boldsymbol{S}_{n-k}} \tag{450}$$

where $\boldsymbol{h}_N = (h_1, \cdots, h_N)^T$ is a constant vector and h_0 a constant offset

- Variance σ_U^2 of prediction residual only depends on $\boldsymbol{h}_N$

$$\begin{aligned} \sigma_U^2(h_0, \boldsymbol{h}_N) &= E\Big\{ \big(U_n - E\{U_n\} \big)^2 \Big\} \\ &= E\bigg\{ \Big(S_n - h_0 - \boldsymbol{h}_N^T \boldsymbol{S}_{n-k} - E\Big\{ S_n - h_0 - \boldsymbol{h}_N^T \boldsymbol{S}_{n-k} \Big\} \Big)^2 \bigg\} \\ &= E\Big\{ \big(S_n - E\{S_n\} - \boldsymbol{h}_N^T \big(\boldsymbol{S}_{n-k} - E\{\boldsymbol{S}_{n-k}\} \big) \big)^2 \Big\} \end{aligned} \tag{451}$$

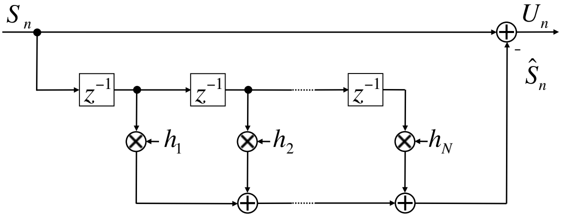
- Mean squared prediction error

$$\begin{aligned} \epsilon_U^2(h_0, \boldsymbol{h}_N) &= \sigma_U^2(\boldsymbol{h}_N) + \mu_U^2(h_0, \boldsymbol{h}_N) = \sigma_U^2(\boldsymbol{h}_N) + E\Big\{ S_n - h_0 - \boldsymbol{h}_N^T \, \boldsymbol{S}_{n-k} \Big\}^2 \\ &= \sigma_U^2(\boldsymbol{h}_N) + \big(\mu_S (1 - \boldsymbol{h}_N^T \boldsymbol{e}_N) - h_0 \big)^2 \end{aligned} \tag{452}$$

- Minimize mean squared prediction error by setting

$$h_0^* = \mu_S \, (1 - \, \boldsymbol{h}_N^T \, \boldsymbol{e}_N) \tag{453}$$

Linear Prediction for Zero-Mean Processes



- The function used for prediction is linear, of the form

$$\boxed{\hat{S}_n = h_1 \cdot S_{n-1} + h_2 \cdot S_{n-2} + \cdots + h_N \cdot S_{n-N} = \boldsymbol{h}_N^T \boldsymbol{S}_{n-1}} \qquad (454)$$

- Mean squared prediction error (same as variance for zero mean)

$$\begin{aligned} \sigma_U^2(\boldsymbol{h}_N) &= E\Big\{(S_n - \hat{S}_n)^2\Big\} = E\Big\{(S_n - \boldsymbol{h}_N^T\boldsymbol{S}_{n-1})(S_n - \boldsymbol{S}_{n-1}^T\boldsymbol{h}_N)\Big\} \\ &= E\{S_n^2\} - 2E\Big\{\boldsymbol{h}_N^T\boldsymbol{S}_{n-1}S_n\Big\} + E\Big\{\boldsymbol{h}_N^T\boldsymbol{S}_{n-1}\boldsymbol{S}_{n-1}^T\boldsymbol{h}_N\Big\} \\ &= E\{S_n^2\} - 2\boldsymbol{h}_N^T E\{S_n\boldsymbol{S}_{n-1}\} + \boldsymbol{h}_N^T E\Big\{\boldsymbol{S}_{n-1}\boldsymbol{S}_{n-1}^T\Big\}\boldsymbol{h}_N \qquad (455) \end{aligned}$$

Autocovariance Matrix and Autocovariance Vector

- Variance $\sigma_S^2 = E\left\{S_n^2\right\}$
- Autocovariance vector (for zero mean: Autocorrelation vector)

$$\boldsymbol{c}_k = E\{S_n \boldsymbol{S}_{n-k}\} = \sigma_S^2 \cdot \begin{pmatrix} \rho_k \\ \vdots \\ \rho_i \\ \vdots \\ \rho_N + k - 1 \end{pmatrix} \quad \text{with} \quad \rho_i = E\{S_n \cdot S_{n-i}\} / \sigma_S^2 \tag{456}$$

- Autocovariance matrix (for zero mean: Autocorrelation matrix)

$$\boldsymbol{C}_N = E\left\{\boldsymbol{S}_{n-1}\boldsymbol{S}_{n-1}^T\right\} = \sigma_S^2 \cdot \begin{pmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{N-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{N-2} \\ \rho_2 & \rho_1 & 1 & \cdots & \rho_{N-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{N-1} & \rho_{N-2} & \rho_{N-3} & \cdots & 1 \end{pmatrix} \tag{457}$$

Optimal Linear Prediction

- Prediction error variance

$$\boxed{\sigma_U^2(\boldsymbol{h}_N) = \sigma_S^2 - 2\boldsymbol{h}_N^T\boldsymbol{c}_k + \boldsymbol{h}_N^T\boldsymbol{C}_N\,\boldsymbol{h}_N} \tag{458}$$

- Minimization of $\sigma_U^2(\boldsymbol{h}_N)$ yields a system of linear equations

$$\boxed{\boldsymbol{C}_N \cdot \boldsymbol{h}_N = \boldsymbol{c}_k} \tag{459}$$

- When $\boldsymbol{C}_N$ is non-singular

$$\boxed{\boldsymbol{h}_N^* = \boldsymbol{C}_N^{-1} \cdot \boldsymbol{c}_k} \tag{460}$$

- Minimum of $\sigma_U^2(\boldsymbol{h}_N)$ is given as (with $(\boldsymbol{C}_N^{-1}\boldsymbol{c}_k)^T = \boldsymbol{c}_k^T\boldsymbol{C}_N^{-1}$)

$$\begin{aligned} \sigma_U^2(\boldsymbol{h}_N^*) &= \sigma_S^2 - 2\,(\boldsymbol{h}_N^*)^T\boldsymbol{c}_k + (\boldsymbol{h}_N^*)^T\boldsymbol{C}_N\,\boldsymbol{h}_N^* \\ &= \sigma_S^2 - 2\left(\boldsymbol{c}_k^T\boldsymbol{C}_N^{-1}\right)\boldsymbol{c}_k + \left(\boldsymbol{c}_k^T\boldsymbol{C}_N^{-1}\right)\boldsymbol{C}_N\left(\boldsymbol{C}_N^{-1}\boldsymbol{c}_k\right) \\ &= \sigma_S^2 - 2\,\boldsymbol{c}_k^T\boldsymbol{C}_N^{-1}\boldsymbol{c}_k + \boldsymbol{c}_k^T\boldsymbol{C}_N^{-1}\boldsymbol{c}_k \\ &= \sigma_S^2 - \boldsymbol{c}_k^T\boldsymbol{C}_N^{-1}\boldsymbol{c}_k = \sigma_S^2 - \boldsymbol{c}_k^T\boldsymbol{h}_N^* \end{aligned} \tag{461}$$

$\Longrightarrow$ Optimal prediction: Signal variance σ_S^2 is reduced by $\boldsymbol{c}_k^T\boldsymbol{C}_N^{-1}\boldsymbol{c}_k = \boldsymbol{c}_k^T\boldsymbol{h}_N^*$

Verification of Optimality

- The optimality of the solution can be verified by inserting $\boldsymbol{h}_N = \boldsymbol{h}_N^* + \boldsymbol{\delta}_N$ into

$$\sigma_U^2(\boldsymbol{h}_N) = \sigma_S^2 - 2\,\boldsymbol{h}_N^T \boldsymbol{c}_k + \boldsymbol{h}_N^T \boldsymbol{C}_N\,\boldsymbol{h}_N \qquad (462)$$

yielding

$$\begin{aligned}
\sigma_U^2(\boldsymbol{h}_N) &= \sigma_S^2 - 2(\boldsymbol{h}_N^* + \boldsymbol{\delta}_N)^T \boldsymbol{c}_k + (\boldsymbol{h}_N^* + \boldsymbol{\delta}_N)^T \boldsymbol{C}_N\,(\boldsymbol{h}_N^* + \boldsymbol{\delta}_N) \\
&= \sigma_S^2 - 2\,(\boldsymbol{h}_N^*)^T \boldsymbol{c}_k - 2\,\boldsymbol{\delta}_N^T \boldsymbol{c}_k + \\
&\qquad (\boldsymbol{h}_N^*)^T \boldsymbol{C}_N\,\boldsymbol{h}_N^* + (\boldsymbol{h}_N^*)^T \boldsymbol{C}_N\,\boldsymbol{\delta}_N + \boldsymbol{\delta}_N^T \boldsymbol{C}_N\,\boldsymbol{h}_N^* + \boldsymbol{\delta}_N^T \boldsymbol{C}_N\,\boldsymbol{\delta}_N \\
&= \sigma_U^2(\boldsymbol{h}_N^*) - 2\boldsymbol{\delta}_N^T \boldsymbol{c}_k + 2\boldsymbol{\delta}_N^T \boldsymbol{C}_N\,\boldsymbol{h}_N^* + \boldsymbol{\delta}_N^T \boldsymbol{C}_N \boldsymbol{\delta}_N \\
&= \sigma_U^2(\boldsymbol{h}_N^*) + \boldsymbol{\delta}_N^T \boldsymbol{C}_N\,\boldsymbol{\delta}_N
\end{aligned} \qquad (463)$$

- The additional term is always non-negative

$$\boldsymbol{\delta}_N^T \boldsymbol{C}_N\,\boldsymbol{\delta}_N \geq 0 \qquad (464)$$

- It is equal to 0 if and only if $\boldsymbol{h}_N = \boldsymbol{h}_N^*$

$\Longrightarrow$ $\boldsymbol{h}_N^*$ is the optimal choice for the prediction parameters

The Orthogonality Principle

- Important property for optimal affine predictors

$$\begin{aligned} E\{U_n\, \boldsymbol{S}_{n-k}\} &= E\Big\{\big(S_n - h_0 - \boldsymbol{h}_N^T \boldsymbol{S}_{n-k}\big)\, \boldsymbol{S}_{n-k}\Big\} \\ &= E\{S_n\, \boldsymbol{S}_{n-k}\} - h_0\, E\{\boldsymbol{S}_{n-k}\} - E\Big\{\boldsymbol{S}_{n-k} \boldsymbol{S}_{n-k}^T\Big\}\, \boldsymbol{h}_N \\ &= \boldsymbol{c}_k + \mu_S^2\, \boldsymbol{e}_N - h_0\, \mu_S\, \boldsymbol{e}_N - (\boldsymbol{C}_N + \mu_S^2\, \boldsymbol{e}_N\, \boldsymbol{e}_N^T)\, \boldsymbol{h}_N \\ &= \boldsymbol{c}_k - \boldsymbol{C}_N \boldsymbol{h}_N + \mu_S\, \boldsymbol{e}_N \big(\, \mu_S\, (1 - \boldsymbol{h}_N^T\, \boldsymbol{e}_N) - h_0 \big) \end{aligned} \tag{465}$$

- Inserting the optimal solutions

$$\boldsymbol{h}_N^* = \boldsymbol{C}_N^{-1} \cdot \boldsymbol{c}_k \qquad \text{and} \qquad h_0^* = \mu_S\, (1 - \boldsymbol{h}_N^T\, \boldsymbol{e}_N) \tag{466}$$

yields

$$\boxed{E\{U_n\, \boldsymbol{S}_{n-k}\} = \boldsymbol{0}} \tag{467}$$

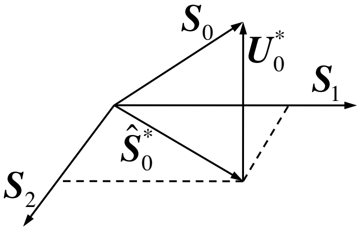
$\Longrightarrow$ For optimal affine prediction, the correlation between the observation vector and the prediction residual is zero

Geometric Interpretation of Orthogonality Principle

- For optimal affine prediction, the correlation between the prediction residual U_n and the observation vector $\boldsymbol{S}_{n-k}$ is zero

$$\boxed{E\{U_n\,\boldsymbol{S}_{n-k}\} = \boldsymbol{0}} \tag{468}$$

- For optimum affine filter design, prediction error should be orthogonal to input signal



- Approximate a vector $\boldsymbol{S}_0$ by a linear combination of $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$
- Best approximation $\hat{\boldsymbol{S}}_0^*$ is given by projection of $\boldsymbol{S}_0$ onto the plane spanned by $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$
- Error vector $\boldsymbol{U}_0^*$ has minimum length and is orthogonal to the projection

One-Step Prediction

- Random variable S_n is predicted using the N directly preceding random variables $\boldsymbol{S}_{n-1} = (S_{n-1}, \cdots, S_{n-N})^T$

- Using $\phi_k = E\{\big(S_n - E\{S_n\}\big)\big(S_{n+k} - E\{S_{n+k}\}\big)\}$, the normal equations are given as

$$\begin{bmatrix} \phi_0 & \phi_1 & \cdots & \phi_{N-1} \\ \phi_1 & \phi_0 & \cdots & \phi_{N-2} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{N-1} & \phi_{N-2} & \cdots & \phi_0 \end{bmatrix} \begin{bmatrix} h_1^N \\ h_2^N \\ \vdots \\ h_N^N \end{bmatrix} = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_N \end{bmatrix} \qquad (469)$$

  where h_k^N represent elements of $\boldsymbol{h}_N^* = (h_1^N, \cdots, h_N^N)^T$

- Changing the equation to

$$\begin{bmatrix} \phi_1 & \phi_0 & \phi_1 & \cdots & \phi_{N-1} \\ \phi_2 & \phi_1 & \phi_0 & \cdots & \phi_{N-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi_N & \phi_{N-1} & \phi_{N-2} & \cdots & \phi_0 \end{bmatrix} \begin{bmatrix} 1 \\ -h_1^N \\ -h_2^N \\ \vdots \\ -h_N^N \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad (470)$$

One-Step Prediction

- Including the prediction error variance for optimal linear prediction using the N preceding samples

$$\begin{aligned} \sigma_N^2 &= \sigma_S^2 - \boldsymbol{c}_1^T \boldsymbol{C}_N^{-1} \boldsymbol{c}_1 = \sigma_S^2 - \boldsymbol{c}_1^T \boldsymbol{h}_N^* \\ &= \phi_0 - h_1^N \phi_1 - h_2^N \phi_2 - \cdots - h_N^N \phi_N \end{aligned} \quad (471)$$

yields and additional row in the matrix

$$\underbrace{\begin{bmatrix} \phi_0 & \phi_1 & \phi_2 & \cdots & \phi_N \\ \phi_1 & \phi_0 & \phi_1 & \cdots & \phi_{N-1} \\ \phi_2 & \phi_1 & \phi_0 & \cdots & \phi_{N-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi_N & \phi_{N-1} & \phi_{N-2} & \cdots & \phi_0 \end{bmatrix}}_{\boldsymbol{C}_{N+1}} \underbrace{\begin{bmatrix} 1 \\ -h_1^N \\ -h_2^N \\ \vdots \\ -h_N^N \end{bmatrix}}_{\boldsymbol{a}_N} = \begin{bmatrix} \sigma_N^2 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (472)$$

- The resulting equation is called **augmented normal equation**

One-Step Prediction

- Multiplying both sides of the augmented normal equation with $\boldsymbol{a}_N^T$ yields

$$\sigma_N^2 = \boldsymbol{a}_N^T \, \boldsymbol{C}_{N+1} \, \boldsymbol{a}_N \tag{473}$$

- Combing the equations for 0 to N preceding samples into one matrix equation yields

$$\boldsymbol{C}_{N+1} \cdot \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ -h_1^N & 1 & \ddots & 0 & 0 \\ -h_2^N & -h_1^{N-1} & \ddots & 0 & 0 \\ \vdots & \vdots & \ddots & 1 & 0 \\ -h_N^N & -h_{N-1}^{N-1} & \cdots & -h_1^1 & 1 \end{bmatrix} = \begin{bmatrix} \sigma_N^2 & X & \cdots & X & X \\ 0 & \sigma_{N-1}^2 & \ddots & X & X \\ 0 & 0 & \ddots & X & X \\ \vdots & \vdots & \ddots & \sigma_1^2 & X \\ 0 & 0 & \cdots & 0 & \sigma_0^2 \end{bmatrix}$$

- Taking the determinant of both sides of the equation gives

$$|\boldsymbol{C}_{N+1}| = \sigma_N^2 \cdot \sigma_{N-1}^2 \cdot \ldots \cdot \sigma_0^2 \tag{474}$$

- Prediction error variance σ_N^2 for optimal linear prediction using the N preceding samples

$$\sigma_N^2 = \frac{|\boldsymbol{C}_{N+1}|}{|\boldsymbol{C}_N|} \tag{475}$$

One-Step Prediction for Autoregressive Processes

- Recall: AR(m) process with mean μ_S and $\boldsymbol{a}_m = (a_1, \cdots, a_m)^T$

$$S_n = Z_n + \mu_S(1 - \boldsymbol{a}_m^T \boldsymbol{e}_m) + \boldsymbol{a}_m^T \boldsymbol{S}_{n-1}^{(m)} \tag{476}$$

- Prediction using N preceding samples in $\boldsymbol{h}_N$ with $N \geq m$:
  Define $\boldsymbol{a}_N = (a_1, \cdots, a_m, 0, \cdots, 0)^T$
- Prediction error

$$U_n = S_n - \boldsymbol{h}_N^T \boldsymbol{S}_{n-1} = Z_n + \mu_S(1 - \boldsymbol{a}_N^T \boldsymbol{e}_N) + (\boldsymbol{a}_N - \boldsymbol{h}_N)^T \boldsymbol{S}_{n-1} \tag{477}$$

- Subtracting the mean $E\{U_n\} = \mu_S(1 - \boldsymbol{a}_N^T \boldsymbol{e}_N) + (\boldsymbol{a}_N - \boldsymbol{h}_N)^T \, E\{\boldsymbol{S}_{n-1}\}$

$$U_n - E\{U_n\} = Z_n + (\boldsymbol{a}_N - \boldsymbol{h}_N)^T \big(\boldsymbol{S}_{n-1} - E\{\boldsymbol{S}_{n-1}\}\big) \tag{478}$$

- Optimal prediction: covariances between U_n and $\boldsymbol{S}_{n-1}$ must be equal to 0

$$\begin{aligned} \boldsymbol{0} &= E\Big\{\big(U_n - E\{U_n\}\big)\big(\boldsymbol{S}_{n-k} - E\{\boldsymbol{S}_{n-k}\}\big)\Big\} \\ &= E\Big\{Z_n\big(\boldsymbol{S}_{n-k} - E\{\boldsymbol{S}_{n-k}\}\big)\Big\} + \boldsymbol{C}_N(\boldsymbol{a}_N - \boldsymbol{h}_N) \end{aligned} \tag{479}$$

yields

$$\boxed{\boldsymbol{h}_N^* = \boldsymbol{a}_N} \tag{480}$$

One-Step Prediction in Gauss-Markov Processes

- Gauss-Markov process is a particular AR(1) process

$$S_n = Z_n + \mu_S(1-\rho) + \rho \cdot S_{n-1}, \tag{481}$$

for which the iid process $\{Z_n\}$ has a Gaussian distribution

- Auto-covariance matrix and its inverse

$$\boldsymbol{C}_2 = \sigma_S^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \qquad \boldsymbol{C}_2^{-1} = \frac{1}{\sigma_S^2(1-\rho^2)} \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix} \tag{482}$$

- Auto-covariance vector

$$\boldsymbol{c}_1 = \sigma_S^2 \begin{pmatrix} \rho \\ \rho^2 \end{pmatrix} \tag{483}$$

- Optimum predictor $\boldsymbol{h}_2^* = \boldsymbol{C}_2^{-1}\boldsymbol{c}_1$

$$\boldsymbol{h}_2^* = \frac{1}{1-\rho^2} \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix} \begin{pmatrix} \rho \\ \rho^2 \end{pmatrix} = \frac{1}{1-\rho^2} \begin{pmatrix} \rho-\rho^3 \\ -\rho^2+\rho^2 \end{pmatrix} = \begin{pmatrix} \rho \\ 0 \end{pmatrix}$$

- First element of $\boldsymbol{h}_N^*$ is equal to ρ, all other elements are equal to 0

One-Step Prediction in Gauss-Markov Processes

- Minimum prediction residual

$$\sigma_U^2 = \frac{|\boldsymbol{C}_2|}{|\boldsymbol{C}_1|} = \frac{\sigma_S^4 - \sigma_S^4\,\rho^2}{\sigma_S^2} = \sigma_S^2\,(1-\rho^2) \qquad (484)$$

- Prediction residual for filter h_1

$$U_n = S_n - h_1 S_{n-1}$$

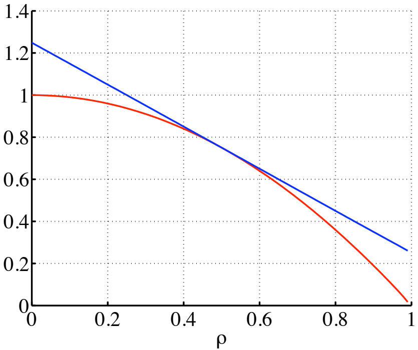
- Average squared error

$$\begin{aligned}\sigma_U^2(h_1) &= E\{U_n^2\} \\ &= \sigma_S^2(1 + h_1^2 - 2\rho h_1)\end{aligned}$$

- Note: Setting derivative to zero

$$\frac{\partial \sigma_U^2(h_1)}{\partial h_1} = \sigma_S^2(2h_1 - 2\rho) \stackrel{!}{=} 0$$

also yields the result $h_1 = \rho$



Prediction Gain

- Prediction gain using $\boldsymbol{\Phi}_N = \boldsymbol{C}_N/\sigma_S^2$ and $\boldsymbol{\phi}_1 = \boldsymbol{c}_1/\sigma_S^2$

$$G_P = \frac{E\{S_n^2\}}{E\{U_n^2\}} = \frac{\sigma_S^2}{\sigma_U^2} = \frac{\sigma_S^2}{\sigma_S^2 - \boldsymbol{c}_1^T\,\boldsymbol{C}_N^{-1}\,\boldsymbol{c}_1} = \frac{1}{1 - \boldsymbol{\phi}_1^T\,\boldsymbol{\Phi}_N^{-1}\,\boldsymbol{\phi}_1}, \qquad (485)$$

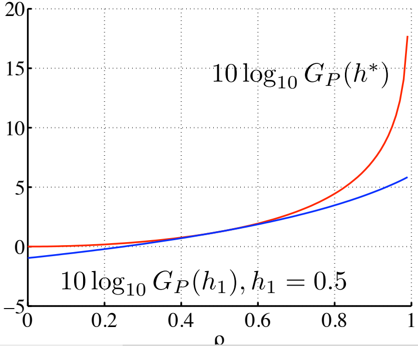
- Prediction gain for optimal prediction in first-order Gauss-Markov process

$$G_P(h^*) = \frac{1}{1-\rho^2} \qquad (486)$$

- Prediction gain for filter h_1

$$G_P(h_1) = \frac{\sigma_S^2}{\sigma_S^2(1 + h_1^2 - 2\rho h_1)} = \frac{1}{1 + h_1^2 - 2\rho h_1}$$

- At high bit rates, $10\log_{10} G_P$: SNR improvement achieved by predictive coding



Optimum Linear One-Step Prediction for $K = 2$

- The normalized auto-correlation matrix and its inverse follow as

$$\boldsymbol{\Phi}_2 = \begin{pmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{pmatrix} \qquad \boldsymbol{\Phi}_2^{-1} = \frac{1}{1-\rho_1^2} \begin{pmatrix} 1 & -\rho_1 \\ -\rho_1 & 1 \end{pmatrix} \tag{487}$$

- With normalized correlation vector

$$\boldsymbol{\phi}_1 = \begin{pmatrix} \rho_1 \\ \rho_2 \end{pmatrix} \tag{488}$$

  we obtain the optimum predictor

$$\begin{aligned} \boldsymbol{h}_2^* &= \boldsymbol{\Phi}_2^{-1}\boldsymbol{\phi}_1 = \frac{1}{1-\rho_1^2} \begin{pmatrix} 1 & -\rho_1 \\ -\rho_1 & 1 \end{pmatrix} \begin{pmatrix} \rho_1 \\ \rho_2 \end{pmatrix} = \frac{1}{1-\rho_1^2} \begin{pmatrix} \rho_1 - \rho_1\rho_2 \\ -\rho_1^2 + \rho_2 \end{pmatrix} \\ &= \frac{1}{1-\rho_1^2} \begin{pmatrix} \rho_1(1-\rho_2) \\ \rho_2 - \rho_1^2 \end{pmatrix} \end{aligned} \tag{489}$$

- For AR(1) sources, where we have $\rho_2 = \rho_1^2$, second coefficient does not improve prediction gain
- General: For AR(m) sources, only m coefficients are unequal to zero

Prediction in Images: Intra-Picture Prediction

- Random variables are samples within that image
- Derivations on linear prediction for zero-mean random variables (subtract μ_S or roughly 127 from 8-bit picture)
- Pictures are typically scanned line-by-line from upper left corner to lower right corner
- 1-d horizontal prediction:

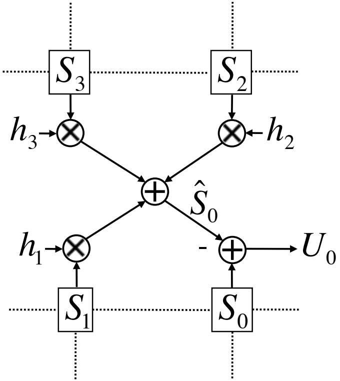
$$\hat{S}_0 = h_1 \cdot S_1$$

- 1-d vertical prediction:

$$\hat{S}_0 = h_2 \cdot S_2$$

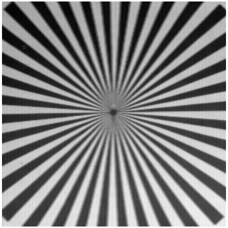
- 2-d prediction:

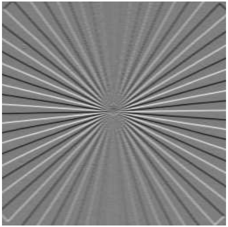
$$\hat{S}_0 = \sum_{i=1}^{3} h_i S_i$$



Prediction Example: Test Pattern

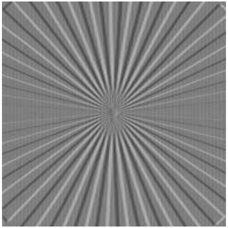
$\sigma_S^2 = 4925.81$
$(s - 127)$

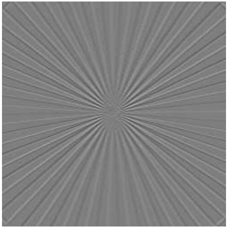




vertical predictor
$h_1 = 0$
$h_2 = 0.932$
$h_3 = 0$
$\sigma_U^2(\boldsymbol{h}) = 646.67$
$G_P = 8.82$ dB

horizontal predictor
$h_1 = 0.953$
$h_2 = 0$
$h_3 = 0$
$\sigma_U^2(\boldsymbol{h}) = 456.17$
$G_P = 10.33$ dB





2-d predictor
$h_1 = 0.911$
$h_2 = 0.871$
$h_3 = -0.788$
$\sigma_U^2(\boldsymbol{h}) = 109.90$
$G_P = 16.51$ dB

Prediction Example: Picture "Lena"

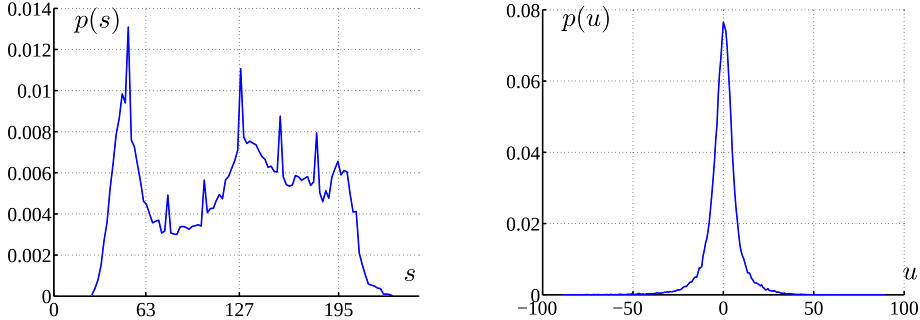
256×256 center cropped picture
$\sigma_S^2 = 2746.43$
$(s - 127)$

vertical predictor
$h_1 = 0$
$h_2 = 0.977$
$h_3 = 0$
$\sigma_U^2(\boldsymbol{h}) = 123.61$
$G_P = 13.47$ dB

horizontal predictor
$h_1 = 0.962$
$h_2 = 0$
$h_3 = 0$
$\sigma_U^2(\boldsymbol{h}) = 212.36$
$G_P = 11.12$ dB

2-d predictor
$h_1 = 0.623$
$h_2 = 0.835$
$h_3 = -0.48$
$\sigma_U^2(\boldsymbol{h}) = 80.35$
$G_P = 15.34$ dB

Prediction Example: PMFs for Picture Lena



- Pmfs $p(s)$ and $p(u)$ change significantly due to prediction operation
- Entropy changes significantly
  (rounding prediction signal towards integer: $E\{\lfloor U_n + 0.5 \rfloor^2\} = 80.47$)

$$H(S) = 7.44\,\text{bit/sample} \qquad H(U) = 4.97\,\text{bit/sample} \tag{490}$$

- Linear prediction can be used for lossless coding: JPEG-LS

Asymptotic Prediction Gain

Consider upper bound for prediction gain: $N \to \infty$

- One-step prediction of a random variable S_n given the countably infinite set of preceding random variables $\{S_{n-1}, S_{n-2}, \cdots\}$ and $\{h_0, h_1, \cdots\}$

$$U_n = S_n - h_0 - \sum_{i=1}^{\infty} h_i \, S_{n-i}, \qquad (491)$$

- Orthogonality criterion: U_n is uncorrelated with all S_{n-k} for $k > 0$
- Furthermore, U_{n-k} for $k > 0$ is fully determined by a linear combination of past input values S_{n-k-i} for $i \geq 0$
- Hence, U_n is uncorrelated with U_{n-k} for $k > 0$

$$\phi_{UU}(k) = \sigma^2_{U,\infty} \cdot \delta(k) \qquad \Longleftrightarrow \qquad \Phi_{UU}(\omega) = \sigma^2_{U,\infty} \qquad (492)$$

where $\sigma^2_{U,\infty}$ is the asymptotic one-step prediction error variance for $N \to \infty$

Asymptotic Prediction Error Variance

- For one-step prediction we showed

$$|\boldsymbol{C}_N| = \sigma^2_{N-1} \cdot \sigma^2_{N-2} \cdot \sigma^2_{N-3} \cdots \sigma^2_0 \tag{493}$$

which yields

$$\frac{1}{N} \ln |\boldsymbol{C}_N| = \ln |\boldsymbol{C}_N|^{\frac{1}{N}} = \frac{1}{N} \sum_{i=0}^{N-1} \ln \sigma_i^2 \tag{494}$$

- If a sequence of numbers $\alpha_0, \alpha_1, \alpha_2, \cdots$ approaches a limit α_∞, the average value approaches the same limit,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=0}^{N-1} \alpha_i = \alpha_\infty \tag{495}$$

- Hence, we can write

$$\lim_{N\to\infty} \ln |\boldsymbol{C}_N|^{\frac{1}{N}} = \lim_{N\to\infty} \frac{1}{N} \sum_{i=0}^{N-1} \ln \sigma_i^2 = \ln \sigma_\infty^2 \tag{496}$$

yielding

$$\sigma_\infty^2 = \exp\left(\lim_{N\to\infty} \ln |\boldsymbol{C}_N|^{\frac{1}{N}} \right) = \lim_{N\to\infty} |\boldsymbol{C}_N|^{\frac{1}{N}} \tag{497}$$

Asymptotic Prediction Error Variance

- Asymptotic One-Step Prediction Error Variance

$$\sigma^2_{U,\infty} = \lim_{N\to\infty} |\boldsymbol{C}_N|^{\frac{1}{N}} \tag{498}$$

- Determinant of $N \times N$ matrix: Product of its eigenvalues $\xi_i^{(N)}$

$$\lim_{N\to\infty} |\boldsymbol{C}_N|^{\frac{1}{N}} = \lim_{N\to\infty} \left(\prod_{i=0}^{N-1} \xi_i^{(N)} \right)^{\frac{1}{N}} = 2^{\left(\lim\limits_{N\to\infty} \sum\limits_{i=0}^{N-1} \frac{1}{N} \log_2 \, \xi_i^{(N)} \right)} \tag{499}$$

- Apply GRENANDER and SZEGÖ's theorem

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=0}^{N-1} G\left(\xi_i^{(N)}\right) = \frac{1}{2\pi} \int_{-\pi}^{\pi} G\left(\Phi(\omega)\right) \, d\omega \tag{500}$$

- Expression using power spectral density

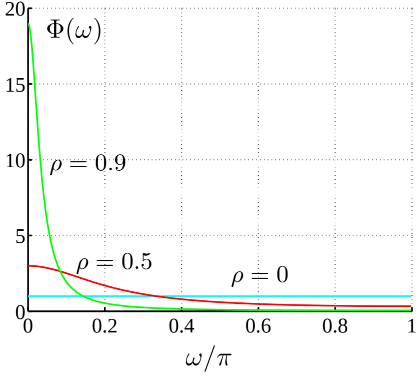
$$\sigma^2_{U,\infty} = \lim_{N\to\infty} |\boldsymbol{C}_N|^{\frac{1}{N}} = 2^{\frac{1}{2\pi} \int_{-\pi}^{\pi} \log_2 \Phi_{SS}(\omega) \, d\omega} \tag{501}$$

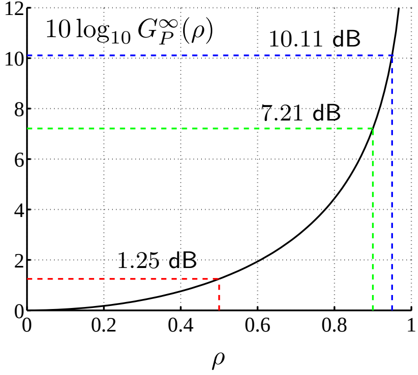
Asymptotic Prediction Gain

- Prediction gain G_P^∞

$$G_P^\infty = \frac{\sigma_S^2}{\sigma_{U,\infty}^2} = \frac{\frac{1}{2\pi}\int_{-\pi}^{\pi}\Phi(\omega)\,d\omega}{2^{\frac{1}{2\pi}\int_{-\pi}^{\pi}\log_2\Phi(\omega)\,d\omega}} \quad \begin{array}{l}\leftarrow \text{Arithmetic mean}\\ \leftarrow \text{Geometric mean}\end{array} \qquad (502)$$

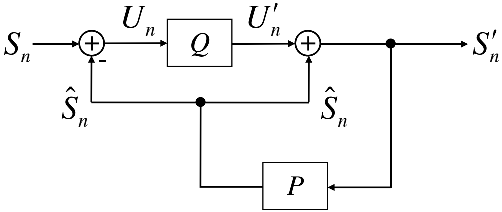
- Result for first-order Gauss-Markov source (can also be computed differently)



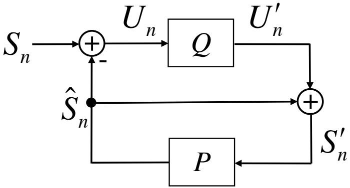


Differential Pulse Code Modulation (DPCM)

- Combining prediction with quantization requires simultaneous reconstruction of predictor at encoder and decoder
  $\Longrightarrow$ Use quantized samples for prediction

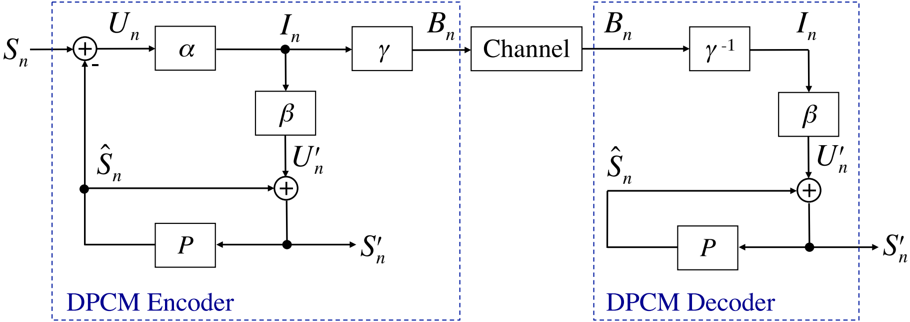


- Re-drawing yields block-diagram with typical DPCM structure



DPCM Codec

- Redrawing with encoder mapping α, lossless coding γ, and decoder mapping β yields DPCM encoder



- DPCM encoder contains DPCM decoder except for γ^{-1}

DPCM and Quantization

- Prediction $\hat{S}_n$ for a sample S_n is generated by linear filtering of reconstructed samples S'_n from the past

$$\hat{S}_n = \sum_{i=1}^{K} h_i \, S'_{n-i} = \sum_{i=1}^{K} h_i \, (S_{n-i} + Q_{n-i}) = \boldsymbol{h}^T \cdot (\boldsymbol{S}_{n-1} + \boldsymbol{Q}_{n-1}) \qquad (503)$$

with $Q_n = S'_n - S_n$ being the quantization error signal

- Prediction error variance (for zero-mean input) is given by

$$\begin{aligned} \sigma_U^2 &= E\{U_n^2\} = E\Big\{(S_n - \hat{S}_n)^2\Big\} = E\Big\{(S_n - \boldsymbol{h}^T\boldsymbol{S}_{n-1} - \boldsymbol{h}^T\boldsymbol{Q}_{n-1})^2\Big\} \\ &= E\{S_n^2\} + \boldsymbol{h}^T E\Big\{\boldsymbol{S}_{n-1}\boldsymbol{S}_{n-1}^T\Big\}\boldsymbol{h} + \boldsymbol{h}^T E\Big\{\boldsymbol{Q}_{n-1}\boldsymbol{Q}_{n-1}^T\Big\}\boldsymbol{h} \qquad (504) \\ &\quad -2\boldsymbol{h}^T E\{S_n\boldsymbol{S}_{n-1}\} - 2\boldsymbol{h}^T E\{S_n\boldsymbol{Q}_{n-1}\} + 2\boldsymbol{h}^T E\Big\{\boldsymbol{S}_{n-1}\boldsymbol{Q}_{n-1}^T\Big\}\boldsymbol{h} \end{aligned}$$

- Defining $\boldsymbol{\Phi} = E\{\boldsymbol{S}_{n-1}\boldsymbol{S}_{n-1}^T\} / \sigma_S^2$ and $\boldsymbol{\phi} = E\{S_n\boldsymbol{S}_{n-1}\} / \sigma_S^2$ we get

$$\begin{aligned} \sigma_U^2 &= \sigma_S^2 \left(1 + \boldsymbol{h}^T\boldsymbol{\Phi}\,\boldsymbol{h} - 2\boldsymbol{h}^T\boldsymbol{\phi}\right) \qquad (505) \\ &\quad +\boldsymbol{h}^T E\Big\{\boldsymbol{Q}_{n-1}\boldsymbol{Q}_{n-1}^T\Big\}\boldsymbol{h} - 2\boldsymbol{h}^T E\{S_n\boldsymbol{Q}_{n-1}\} + 2\boldsymbol{h}^T E\Big\{\boldsymbol{S}_{n-1}\boldsymbol{Q}_{n-1}^T\Big\}\boldsymbol{h} \end{aligned}$$

DPCM for a Gauss-Markov Source

- Calculate $R(D)$ for zero-mean Gauss-Markov process

$$S_n = Z_n + \rho \cdot S_{n-1} \tag{506}$$

- Consider a one-tap linear prediction filter $\boldsymbol{h} = [h]$
- Normalized auto-correlation matrix $\mathbf{\Phi} = [1]$ and cross-correlation $\boldsymbol{\phi} = [\rho]$
- Prediction error variance

$$\begin{aligned}\sigma_U^2 \;=\; & \sigma_S^2 \left(1 + h^2 - 2\,h\,\rho\right) + h^2 E\{Q_{n-1}^2\} \\ & -2hE\{S_n Q_{n-1}\} + 2h^2 E\{S_{n-1} Q_{n-1}\}\end{aligned} \tag{507}$$

- Using $S_n = Z_n + \rho \cdot S_{n-1}$, the second row in above equation becomes

$$\begin{aligned}& -2hE\{S_n Q_{n-1}\} + 2h^2 E\{S_{n-1} Q_{n-1}\} \\ & \;=\; -2hE\{Z_n Q_{n-1}\} - 2h\rho E\{S_{n-1} Q_{n-1}\} + 2h^2 E\{S_{n-1} Q_{n-1}\} \\ & \;=\; -2hE\{Z_n Q_{n-1}\} + 2h(h-\rho) E\{S_{n-1} Q_{n-1}\}\end{aligned} \tag{508}$$

- With setting $h = \rho$, we have

$$E\{Z_n Q_{n-1}\} = 0 \qquad 2h(h-\rho)E\{S_{n-1} Q_{n-1}\} = 0 \tag{509}$$

DPCM for a Gauss-Markov Source

- For $h = \rho$, expression for prediction error variance simplifies to

$$\sigma_U^2 = \sigma_S^2 \left(1 - \rho^2\right) + \rho^2 E\{Q_{n-1}^2\} \tag{510}$$

- Assume: Prediction error for Gaussian source has also Gaussian distribution
- Model expression for quantization error $D = E\{Q_{n-1}^2\}$ by an operational distortion rate function

$$D(R) = \sigma_U^2 \cdot g(R) \tag{511}$$

- Expression for prediction error variance becomes dependent on rate

$$\boxed{\sigma_U^2 = \sigma_S^2 \cdot \frac{1 - \rho^2}{1 - g(R)\,\rho^2}} \tag{512}$$

- Operational distortion-rate function for DPCM of Gauss-Markov

$$\boxed{D(R) = \sigma_U^2 \cdot g(R) = \sigma_S^2 \cdot \frac{1 - \rho^2}{1 - g(R)\,\rho^2} \cdot g(R)} \tag{513}$$

Computation of DPCM Distortion-Rate Function

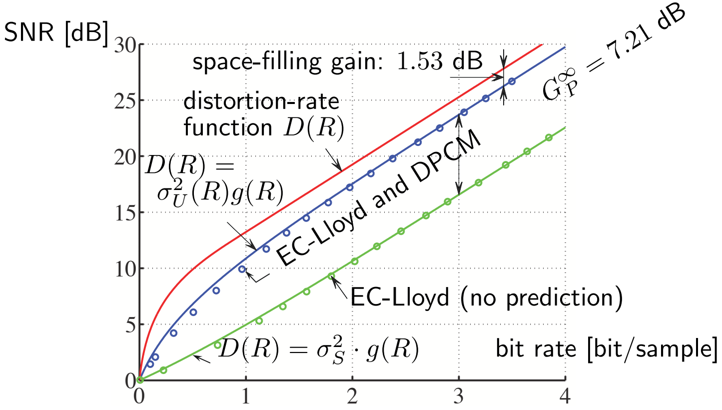
- Operational distortion-rate function for DPCM and ECSQ for a Gauss-Markov source

$$\boxed{D(R) = \sigma_U^2 \cdot g(R) = \sigma_S^2 \cdot \frac{1-\rho^2}{1-g(R)\,\rho^2} \cdot g(R)} \qquad (514)$$

- Algorithm for designing ECSQ inside DPCM loop
  1. Initialization with a small value of λ, set $s'_n = s_n, \forall n$ and $h = \rho$
  2. Create signal u_n using s'_n and DCPM
  3. Design ECSQ (α, β, γ) using signal u_n and the current value of λ by minimizing $D + \lambda R$
  4. Conduct DPCM encoding/decoding using ECSQ (α, β, γ)
  5. Measure $\sigma_U^2(R)$ as well as rate R and distortion D
  6. Increase λ and start again with step 2

- Algorithm shows problems at low bit rates: Instabilities

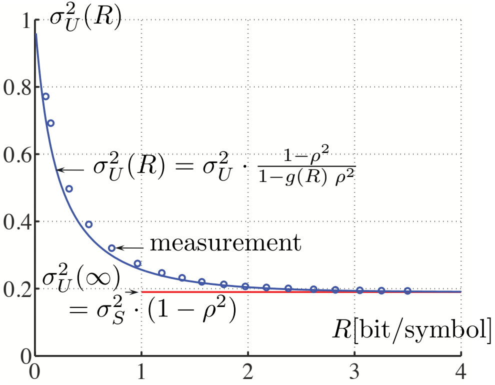
Comparison of Theoretical and Experimental Results



- For high rates and Gauss-Markov sources, shape and memory gain achievable
- Space-filling gain can only be achieved using vector quantization
- Theoretical model provides a useful description

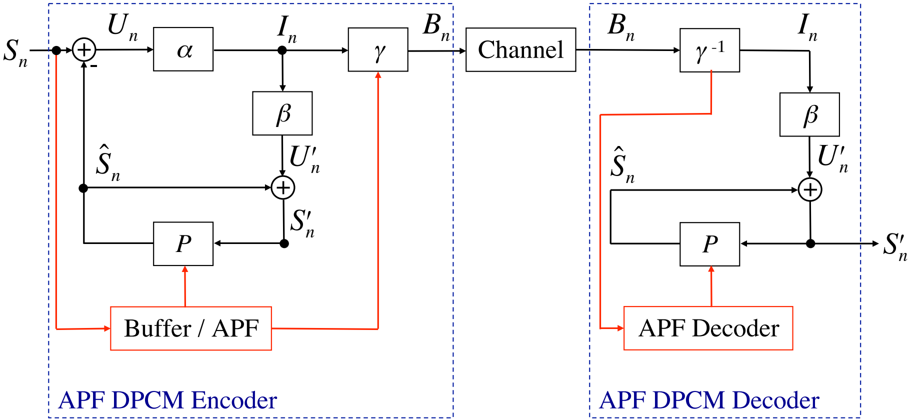
Comparison of Theoretical and Experimental Results

- Prediction error variance σ_U^2 depends on bit rate
- Theoretical model provides a useful description



Adaptive Differential Pulse Code Modulation (ADPCM)

- For quasi-stationary sources like speech, fixed predictor is not well suited
- ADPCM: Adapt the predictor based on the recent signal characteristics
- Forward adaptation: Send new predictor values (requires additional bit rate)



Forward-Adaptive Prediction: Motion Compensation

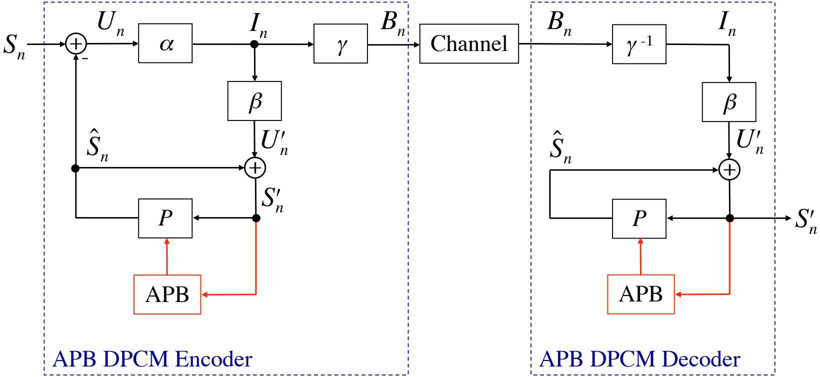
- Since predictor values are sent, extend prediction to vectors/blocks
- Use statistical dependencies between two pictures
- Prediction signal obtained by searching a region in a previously decoded picture that best matches the block to be coded
- Let $s[x,y]$ represent intensity at location (x,y)
- Let $s'[x,y]$ represent intensity in a previously decoded picture at (x,y)

$$J = \min_{(dx,dy)} \sum_{x,y} (s[x,y] - s'[x-dx, y-dy])^2 + \lambda \cdot R(dx,dy) \qquad (515)$$

- Predicted signal is specified through motion vector (dx, dy)
- $R(dx, dy)$ represents the number of bits required for coding the motion vector
- Prediction error $u[x,y]$ is quantized (often using transform coding)
- Bit rate is sum of motion vector and prediction residual bit rate

Backward Adaptive DPCM

- Backward adaptation: Use predictor computed from recently decoded signal
  - No additional bit rate
  - Error resilience issues
  - Accuracy of predictor
  - Decoder complexity

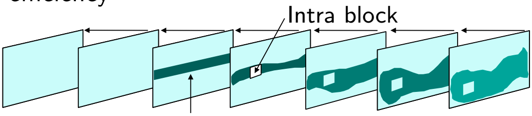


Transmission Errors in DPCM

- When transmission error occurs, DPCM causes error propagation

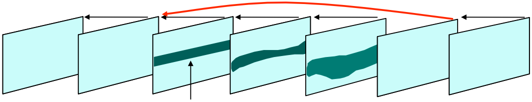
Example: Motion compensation in video coding

- Try to conceal image parts that are in error
- Code lost image parts without referencing concealed image parts helps but reduces coding efficiency



Concealed image part

- Use "clean" reference picture for motion compensation



Concealed image part

Chapter Summary

Prediction

- Estimate random variable from already observed random variables
- Optimal predictor: Conditional mean

Linear and affine prediction

- Simple and efficient structure
- Optimal predictor given by Wiener-Hopf equation
- AR(m) processes: Optimal predictor has m coefficients
- Optimal prediction error is orthogonal to input signal
- Non-matched predictor can increase signal variance

Predictive quantization: DPCM

- Combination of affine prediction and ECSQ is simple and efficient
- Can exploit linear dependencies between samples
- Forward and backward adaptation
- Transmission errors cause error propagation

Exercise 20

Given is a stationary random process $\mathbf{S} = \{S_n\}$.

We consider affine prediction of a random variable S_n given the N preceding random variables $\mathbf{S}_{n-1} = [S_{n-1}\ S_{n-2}\ \cdots\ S_{n-N}]^T$.

Derive all formulas (as given below) as function of the mean μ_s, the variance σ_S^2, the N-th order autocovariance matrix $\mathbf{C}_N$ and the autocovariance vector $\mathbf{c}_1 = E\{(S_n - \mu_S)(\mathbf{S}_{n-1} - \mu_S \mathbf{e}_N)\}$, where $\mathbf{e}_N$ is a N-dimensional vector with all entries equal to 1.

(a) Derive the affine predictor that minimizes the mean squared prediction error.

(b) Derive expressions for the mean and the variance of the resulting prediction error as well as for the mean squared error.

(c) Derive the affine predictor and the resulting mean, variance and mean squared error for the special case $N = 1$, menaing that a random variable S_n is predicted using the random variable S_{n-1}. The correlation coefficient between successive random variables is ρ.

Exercise 21

In image and video coding, a sample S_n is often predicted by directly using a previous sample S_{n-1}, i.e., by $\hat{S}_n = S_{n-1}$.

Consider a zero-mean stationary process $\mathbf{S} = \{S_n\}$ with the first-order correlation factor ρ.

(a) For what correlation factors ρ do we observe a prediction gain (the mean squared prediction error is smaller than the second moment of the input)?

(b) For what correlation factors is the loss versus optimal linear prediction smaller than 0.1 dB?

Exercise 22 - Part I

Consider prediction in images. Assume that an image can be considered as a realization of a stationary 2-d process with mean μ_S and variance σ_S^2.
We want to linearily predict a current sample based on up to three (already coded) neigbouring samples: the sample left of the current sample, the sample above the current sample, and the sample to the top-left of the current sample.
The correlation factor between two horizontally adjacent samples is ρ_H, the correlation factor between two vertically adjacent samples is ρ_V, and the correlation factor between two diagonally adjacent samples is ρ_D (same in both directions).
The goal is to design linear predictors that minimize the mean squared prediction error. The mean μ_S is subtracted before doing the prediction.

(a) Assume that $\rho_H > \rho_V$.
Compare optimal linear prediction using only the horizontally adjacent sample and optimal linear prediction using both the horizontally and the vertically adjacent sample.
Under which cicumstances is the prediction using both samples better than the prediction using only the horizontally adjacent sample?

Exercise 22 - Part II

(b) Consider the special case $\rho_H = \rho_V = \rho$ and $\rho_D = \rho^2$.
Derive the prediction gain $g = \sigma_S^2 / \varepsilon^2$ for the optimal vertical predictors using

- the sample to the left
- the sample to the left and the sample above
- the sample to the left, the sample above, and the sample to the top-left

What are the prediction gains in dB for $\rho = 0.95$?

Exercise 23 – Part I

Given is a stationary AR(2) process

$$S_n = Z_n + \alpha_1 \cdot S_{n-1} + \alpha_2 \cdot S_{n-2}$$

where $\{Z_n\}$ represents zero-mean white noise.
The AR parameters are $\alpha_1 = 0.7$ and $\alpha_2 = 0.2$.

(a) Determine the correlation factors ρ_1 and ρ_2, where ρ_1 is the correlation factor between adjacent samples S_n and S_{n-1}, and ρ_2 is the correlation factor between samples S_n and S_{n-2} that are two sampling intervals apart.

(b) Derive the optimal linear predictor (minimizing the MSE) using the 2 previous samples.
Determine the prediction gain in dB.

(c) Derive the optimal linear predictor (minimizing the MSE) using only the directly preceeding sample.
What is the prediction gain in dB?.
What is the loss relative to an optimal prediction using the last two samples?

Exercise 23 – Part II

(d) Can the linear predictor using the directly preceeding sample, given by

$$U_n = S_n - \rho_1 \cdot S_{n-1}.$$

be improved by adding a second prediction stage

$$V_n = U_n - h \cdot U_{n-1}?$$

What is the optimal linear predictor for the second prediction stage?
What is the prediction gain achieved by the second prediction stage?
How big is the loss versus optimal linear prediction using the last two samples?

Exercise 24

Consider a zero-mean Gauss-Markov process with the correlation factor $\rho = 0.9$. The Gauss-Markov source is coded using DPCM at high rates. The quantizer is an entropy-contrained Lloyd quantizer with optimal entropy coding.

(a) Neglect the quantization and derive the optimal linear predictor (minimizing the MSE) using the previous sample.
Determine the prediction gain.

(b) Use the predictor derived in (a) inside the DPCM loop.
Assume that the prediction error has a Gaussian distribution.
What is the approximate coding gain compared to ECSQ without prediction at the rates $R_1 = 1$ bit per sample, $R_2 = 2$ bit per sample, $R_3 = 3$ bit per sample, $R_4 = 4$ bit per sample, and $R_5 = 8$ bit per sample?