Exercises with solutions (Set F)

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 requested 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 an N-dimensional vector with all entries equal to 1.

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

    Solution:
    An affine predictor can be written as

    $$\hat{S_n} = h_0 + \mathbf{h}_N^T \cdot \mathbf{S}_{n-k}$$

    with h_0 and

    $$\mathbf{h}_N = [h_1\, h_2\, \cdots\, h_N]^T$$

    being the parameters of the affine predictor.
    The prediction error is then given by

    $$U_n = S_n - \hat{S_n} = S_n - h_0 - \mathbf{h}_N^T\, \mathbf{S}_{n-k}$$

    For the mean of the prediction error, we obtain

    $$\begin{aligned} \mu_U &= E\{U_n\} = E\{S_n\} - h_0 - \mathbf{h}_N^T\, E\{\mathbf{S}_{n-k}\} \\ &= \mu_S - h_0 - \mu_S\, \mathbf{h}_N^T\, \mathbf{e}_N \\ &= \mu_S\, (1 - \mathbf{h}_N^T\, \mathbf{e}_N) - h_0 \end{aligned}$$

    The variance of the prediction error is given by

    $$\begin{aligned} \sigma_U^2 &= E\left\{\big(U - E\{U\}\big)^2\right\} \\ &= E\left\{\big(S_n - h_0 - \mathbf{h}_N^T\, \mathbf{S}_{n-k} - \mu_S\, (1 - \mathbf{h}_N^T\, \mathbf{e}_N) + h_0\big)^2\right\} \\ &= E\left\{\big((S_n - \mu_S) - \mathbf{h}_N^T\, (\mathbf{S}_{n-k} - \mu_S \mathbf{e}_N)\big)^2\right\} \\ &= E\left\{\big(S_n - \mu_S\big)^2\right\} - 2\, \mathbf{h}_N^T\; E\left\{\big(S_n - \mu_S\big)\big(\mathbf{S}_{n-k} - \mu_S \mathbf{e}_N\big)\right\} \\ &\quad + E\left\{\big(\mathbf{S}_{n-k} - \mu_S \mathbf{e}_N\big)\big(\mathbf{S}_{n-k} - \mu_S \mathbf{e}_N\big)^T\right\} \\ &= \sigma_S^2 - 2\, \mathbf{h}_N^T\, \mathbf{c}_1 + \mathbf{h}_N^T\, \mathbf{C}_N\, \mathbf{h}_N \end{aligned}$$

    The mean squared prediction error can be written as

    $$\begin{aligned} \varepsilon_U^2 &= \sigma_U^2 + \mu_U^2 \\ &= \sigma_S^2 - 2\, \mathbf{h}_N^T\, \mathbf{c}_1 + \mathbf{h}_N^T\, \mathbf{C}_N\, \mathbf{h}_N + \Big(\mu_S\, (1 - \mathbf{h}_N^T\, \mathbf{e}_N) - h_0\Big)^2 \end{aligned}$$

    The MSE is a convex function of the parameters h_k, with $0 \le k \le N$.

We first consider the minimization of the MSE with respect to h_0. Since the first term of the MSE (which represents the variance) is independent of h_0, the minimization yields

$$\begin{aligned}
\frac{\partial}{\partial h_0}\,\varepsilon_U^2 &= 2\Big(\mu_S\,(1-\mathbf{h}_N^T\,\mathbf{e}_N) - h_0\Big)(-1) \\
0 &= \mu_S\,(1-\mathbf{h}_N^T\,\mathbf{e}_N) - h_0 \\
h_0 &= \mu_S\,(1-\mathbf{h}_N^T\,\mathbf{e}_N) = \mu_S\,\left(1-\sum_{i=1}^{N} h_i\right)
\end{aligned}$$

The resulting mean squared prediction error is

$$\varepsilon_U^2 = \sigma_U^2 = \sigma_S^2 - 2\,\mathbf{h}_N^T\,\mathbf{c}_1 + \mathbf{h}_N^T\,\mathbf{C}_N\,\mathbf{h}_N$$

For deriving the vector $\mathbf{h}_N$ that minimizes the variance, we write the vector and matrix multiplications as sums:

$$\sigma_U^2 = \sigma_S^2 - 2\sum_{i=1}^{N} h_i(\mathbf{c}_1)_i + \sum_{i=0}^{N}\sum_{j=0}^{N} h_i h_j (\mathbf{C}_N)_{ij}$$

Since σ_U^2 is a convex function with respect to the parameters h_k, with $1 \le k \le N$, the parameters that minimize the variance can be found be setting the first derivatives equal to 0.
For $1 \le k \le 1$, we have

$$\begin{aligned}
\frac{\partial}{\partial h_k}\,\sigma_U^2 &= -2\,(\mathbf{c}_1)_k + 2\,h_k\,(\mathbf{C}_N)_{kk} + \sum_{i\neq k} h_i\,(\mathbf{C}_N)_{ik} + \sum_{i\neq k} h_i\,(\mathbf{C}_N)_{ki} \\
0 &= -2\,(\mathbf{c}_1)_k + 2\sum_{i=0}^{N} h_i\,(\mathbf{C}_N)_{ik},
\end{aligned}$$

yielding

$$\begin{aligned}
\sum_{i=0}^{N} (\mathbf{C}_N)_{ki}\,h_i &= (\mathbf{c}_1)_k \\
(\mathbf{C}_N)_k\,\mathbf{h}_N &= (\mathbf{c}_1)_k,
\end{aligned}$$

where $(\mathbf{C}_N)_k$ denotes the k-th row of the matrix $\mathbf{C}_N$.
By combining all N equations, we obtain the matrix equation

$$\mathbf{C}_N \cdot \mathbf{h}_N = \mathbf{c}_1$$

And by multiplication with the inverse autocovariance matrix from the front, we obtain

$$\mathbf{h}_N = \mathbf{C}_N^{-1} \cdot \mathbf{c}_1$$

Hence, the affine predictor that minimizes the mean squared prediction error is given by

$$\begin{aligned}
\mathbf{h}_N &= \mathbf{C}_N^{-1} \cdot \mathbf{c}_1 \\
h_0 &= \mu_S\,(1-\mathbf{h}_N^{\mathrm{T}}\,\mathbf{e}_N) = \mu_S\,(1-\mathbf{c}_1^{\mathrm{T}}\,\mathbf{C}_N^{-1}\,\mathbf{e}_N)
\end{aligned}$$

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

<u>*Solution:*</u>

For the mean of the prediction error, we obtain

$$\begin{aligned}\mu_U &= \mu_S\,(1-\mathbf{h}_N^{\mathrm{T}}\,\mathbf{e}_N) - h_o \\ &= \mu_S\,(1-\mathbf{h}_N^{\mathrm{T}}\,\mathbf{e}_N) - \mu_S\,(1-\mathbf{h}_N^{\mathrm{T}}\,\mathbf{e}_N) \\ &= 0\end{aligned}$$

By inserting $\mathbf{C}_N\mathbf{h}_N = \mathbf{c}_1$ into the expression for the variance, we obtain

$$\begin{aligned}\sigma_U^2 &= \sigma_S^2 - 2\,\mathbf{h}_N^T\,\mathbf{c}_1 + \mathbf{h}_N^T\,\mathbf{C}_N\,\mathbf{h}_N \\ &= \sigma_S^2 - 2\,\mathbf{h}_N^T\,\mathbf{c}_1 + \mathbf{h}_N^T\,\mathbf{c}_1 \\ &= \sigma_S^2 - \mathbf{h}_N^T\,\mathbf{c}_1\end{aligned}$$

and

$$\begin{aligned}\sigma_U^2 &= \sigma_S^2 - \mathbf{h}_N^T\,\mathbf{c}_1 \\ &= \sigma_S^2 - \left(\mathbf{C}_N^{-1}\,\mathbf{c}_1\right)^T\,\mathbf{c}_1 \\ &= \sigma_S^2 - \mathbf{c}_1^T\,\mathbf{C}_N^{-1}\,\mathbf{c}_1\end{aligned}$$

Finally, for the mean squared prediction error, we obtain

$$\begin{aligned}\varepsilon_U^2 &= \sigma_U^2 + \mu_U^2 \\ &= \sigma_U^2 \\ &= \sigma_S^2 - \mathbf{c}_1^T\,\mathbf{C}_N^{-1}\,\mathbf{c}_1\end{aligned}$$

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

<u>*Solution:*</u>

The autocovariance matrix $\mathbf{C}_N$ and the autocovariance vector $\mathbf{c}_1$ are given by

$$\begin{aligned} \mathbf{C}_N &= E\left\{ \left(S_n - \mu_S\right)^2 \right\} = \sigma_S^2 \\ \mathbf{c}_1 &= E\left\{ \left(S_n - \mu_S\right)\left(S_{n-1} - \mu_S\right) \right\} = \rho\, \sigma_S^2 \end{aligned}$$

For the coefficient h_1 of the affine predictor that minimizes the MSE ε_U^2, we obtain

$$\begin{aligned} \mathbf{C}_N\, h_1 &= \mathbf{c}_1 \\ \sigma_S^2\, h_1 &= \rho\, \sigma_S^2 \\ h_1 &= \rho \end{aligned}$$

And for the coefficient h_0, we have

$$\begin{aligned} h_0 &= \mu_S(1 - h_1) \\ &= \mu_S(1 - \rho) \end{aligned}$$

So, the affine predictor is given by

$$\hat{S_n} = \rho \cdot S_{n-1} + \mu_S(1 - \rho)$$

The resulting mean is

$$\mu_U = 0,$$

the variance

$$\sigma_U^2 = \sigma_S^2 - h\, \mathbf{c}_1 = \sigma_S^2 - h\, \rho\, \sigma_S^2 = \sigma_S^2\, (1 - \rho^2),$$

and the MSE

$$\varepsilon_U^2 = \sigma_U^2 + \mu_U^2 = \sigma_S^2\, (1 - \rho^2)$$

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)?

<u>*Solution:*</u>
The MSE (or variance) of the prediction error is given by

$$\begin{aligned} \varepsilon^2_{U,1} &= \sigma^2_S - 2\,h\,\mathbf{c}_1 + h^2\,\mathbf{C}_N \\ &= \sigma^2_S - 2\rho\sigma^2_S + \sigma^2_S \\ &= 2\sigma^2_S\,(1-\rho) \end{aligned}$$

A positive prediction gain is observed for

$$\begin{aligned} \varepsilon^2_{U,1} &< \varepsilon^2_S \\ 2\sigma^2_S\,(1-\rho) &< \sigma^2_S \\ 2(1-\rho) &< 1 \\ (1-\rho) &< \frac{1}{2} \\ \rho &> \frac{1}{2} \end{aligned}$$

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

In optimal prediction, we have $h = \rho$, yielding the MSE

$$\begin{aligned} \varepsilon^2_{U,opt} &= \sigma^2_S - h\,\mathbf{c}_1 + h^2\,\mathbf{C}_N \\ &= \sigma^2_S - 2\rho^2\sigma^2_S + \rho^2\sigma^2_S \\ &= \sigma^2_S\,(1-\rho^2) \end{aligned}$$

The loss in prediction gain is smaller than 0.1 dB, if we have

$$\begin{aligned} -10\log_{10}\frac{\varepsilon^2_{U,opt}}{\varepsilon^2_{U,1}} &< \frac{1}{10} \\ \log_{10}\frac{\sigma^2_S\,(1-\rho^2)}{2\sigma^2_S\,(1-\rho)} &> -\frac{1}{100} \\ \frac{1+\rho}{2} &> 10^{-\frac{1}{100}} \\ \rho &> 2\cdot 10^{-\frac{1}{100}} - 1 \;\approx\; 0.9545 \end{aligned}$$

22. 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 to the 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 from both the current sample and the samples used for prediction 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?

Solution:
We first consider optimal linear prediction using only the horizontally adjacant sample. The predictor is given by

$$\hat{S}_X = \mu_S + h\,(S_H - \mu_S)$$

yielding the prediction error

$$U_H = S_X - \hat{S}_X = S_X - \mu_S - h\,(S_H - \mu_S)$$

For the mean squared prediction error, we obtain

$$\begin{aligned}
\varepsilon_H^2 &= E\left\{U_H^2\right\} = E\left\{\left(S_X - \mu_S + h\,(S_H - \mu_S)\right)^2\right\} \\
&= E\left\{\left(S_X - \mu_S\right)^2\right\} - 2h\,E\left\{\left(S_X - \mu_S\right)\right)\left(S_H - \mu_S\right)\right)\right\} \\
&\quad + h^2\,E\left\{\left(S_H - \mu_S\right)^2\right\} \\
&= \sigma_S^2 - 2h\rho_H\sigma_S^2 + h^2\sigma_S^2 \\
&= \sigma_S^2(1 - 2h\rho_H + h^2)
\end{aligned}$$

The prediction coefficient that minimizes the MSE is given by

$$h = \rho_H$$

yielding

$$\varepsilon_H^2 = \sigma_S^2\,\left(1 - \rho_H^2\right)$$

Now, we consider optimal linear prediction using the horizontally and the vertically adjacant sample. The predictor is given by

$$\hat{S}_X = \mu_S + \mathbf{h}_2\,(\mathbf{S}_{HV} - \mu_S\mathbf{e}_2)$$

with $\mathbf{h}_2 = [h\ v]^T$, $\mathbf{S}_{HV} = [S_H\ S_V]^T$, and $\mathbf{e}_2 = [1\ 1]^T$.
For the mean squared prediction error, we obtain

$$\begin{aligned}
\varepsilon^2_{HV} &= E\left\{U^2_{HV}\right\} = E\left\{\left(S_X - \mu_S + \mathbf{h}_2\left(\mathbf{S}_{HV} - \mu_S \mathbf{e}_2\right)\right)^2\right\} \\
&= E\left\{\left(S_X - \mu_S\right)^2\right\} - 2\mathbf{h}_2\, E\left\{\left(S_X - \mu_S\right)\left(\mathbf{S}_H - \mu_S \mathbf{e}_2\right)\right\} \\
&\quad +\mathbf{h}_2^T\, E\left\{\left(\mathbf{S}_{HV} - \mu_S \mathbf{e}_2\right)\left(\mathbf{S}_{HV} - \mu_S \mathbf{e}_2\right)^T\right\} \\
&= \sigma_S^2 - 2\mathbf{h}_2 \mathbf{c} + \mathbf{h}_2^T \mathbf{C}\, \mathbf{h}_2
\end{aligned}$$

with the autocovariance matrix

$$\mathbf{C} = E\left\{\begin{bmatrix} S_H - \mu_S \\ S_V - \mu_S \end{bmatrix} \cdot \begin{bmatrix} S_H - \mu_S & S_V - \mu_S \end{bmatrix}\right\} = \sigma_S^2 \begin{bmatrix} 1 & \rho_D \\ \rho_D & 1 \end{bmatrix}$$

and the autocovariance vector

$$\mathbf{c} = E\left\{\begin{bmatrix} S_H - \mu_S \\ S_V - \mu_S \end{bmatrix} \cdot \left(S_x - \mu_S\right)\right\} = \sigma_S^2 \begin{bmatrix} \rho_H \\ \rho_V \end{bmatrix}$$

The prediction coefficients that minimize the MSE are given by the solution of

$$\sigma_S^2 \begin{bmatrix} 1 & \rho_D \\ \rho_D & 1 \end{bmatrix} \cdot \begin{bmatrix} h \\ v \end{bmatrix} = \sigma_S^2 \begin{bmatrix} \rho_H \\ \rho_V \end{bmatrix}$$

Hence, we have the two equations

$$\begin{aligned}
h + \rho_D \cdot v &= \rho_H \\
\rho_D \cdot h + v &= \rho_V
\end{aligned}$$

Multiplying the second equation by $-\rho_D$ and adding the result to the first equation yields

$$\begin{aligned}
\left(1 - \rho_D^2\right) h &= \rho_H - \rho_D \rho_V \\
h &= \frac{\rho_H - \rho_D \rho_V}{1 - \rho_D^2}
\end{aligned}$$

Similarly, we obtain

$$v = \frac{\rho_V - \rho_D \rho_H}{1 - \rho_D^2}$$

yieling

$$\mathbf{h}_2 = \frac{1}{1 - \rho_D^2} \begin{bmatrix} \rho_H - \rho_D \rho_V \\ \rho_V - \rho_D \rho_H \end{bmatrix}$$

For the MSE, we obtain

$$\begin{aligned}
\varepsilon^2_{HV} &= \sigma_S^2 - \mathbf{c}^T\, \mathbf{h}_2 \\
&= \sigma_S^2 - \frac{\sigma_S^2}{1 - \rho_D^2} \begin{bmatrix} \rho_H & \rho_V \end{bmatrix} \begin{bmatrix} \rho_H - \rho_D \rho_V \\ \rho_V - \rho_D \rho_H \end{bmatrix} \\
&= \sigma_S^2 - \frac{\sigma_S^2}{1 - \rho_D^2} \left(\rho_H^2 + \rho_V^2 - 2\rho_D \rho_H \rho_V\right) \\
&= \sigma_S^2 \left(1 - \frac{\rho_H^2 + \rho_V^2 - 2\rho_D \rho_H \rho_V}{1 - \rho_D^2}\right)
\end{aligned}$$

The difference between the MSE ε_H^2 and ε_{HV}^2 is

$$\begin{aligned}
\varepsilon_H^2 - \varepsilon_{HV}^2 &= \sigma_S^2 (1 - \rho_H^2) - \sigma_S^2 \left(1 - \frac{\rho_H^2 + \rho_V^2 - 2\rho_D\rho_H\rho_V}{1 - \rho_D^2}\right) \\
&= \sigma_S^2 \left(\frac{\rho_H^2 + \rho_V^2 - 2\rho_D\rho_H\rho_V}{1 - \rho_D^2} - \rho_H^2\right) \\
&= \sigma_S^2 \left(\frac{\rho_H^2 + \rho_V^2 - 2\rho_D\rho_H\rho_V - \rho_H^2 + \rho_D^2\rho_H^2}{1 - \rho_D^2}\right) \\
&= \sigma_S^2 \left(\frac{(\rho_V)^2 - 2(\rho_D\rho_H)(\rho_V) + (\rho_D\rho_H)^2}{1 - \rho_D^2}\right) \\
&= \sigma_S^2 \frac{(\rho_V - \rho_D\rho_H)^2}{1 - \rho_D^2}
\end{aligned}$$

The MSE for the 2-sample prediction is never greater than the MSE for the 1-sample prediction. In general ($\rho_D \neq \frac{\rho_V}{\rho_H}$), the MSE can be reduced by using the vertically adjacent sample in addition to the horizontally adjacent sample for prediction.

(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 $\rho = 0.95$?

Solution:
For the predictor using the horizontally adjacent sample, we have

$$\begin{aligned} g_H &= \frac{\sigma_S^2}{\varepsilon_H^2} \\ &= \frac{\sigma_S^2}{\sigma_S^2\,(1-\rho_H^2)} \\ &= \frac{1}{1-\rho^2} \end{aligned}$$

For the predictor that uses the horizontally and the vertically adjacent sample, we have

$$\begin{aligned} g_{HV} &= \frac{\sigma_S^2}{\varepsilon_{HV}^2} \\ &= \frac{\sigma_S^2}{\sigma_S^2\left(1-\frac{\rho_H^2+\rho_V^2-2\rho_D\rho_H\rho_V}{1-\rho_D^2}\right)} \\ &= \frac{1-\rho_D^2}{1-\rho_D^2-\rho_H^2-\rho_V^2+2\rho_D\rho_H\rho_V} \\ &= \frac{1-\rho^4}{1-\rho^4-\rho^2-\rho^2+2\rho^4} \\ &= \frac{1-\rho^4}{1-2\rho^2+\rho^4} = \frac{\left(1-\rho^2\right)\left(1+\rho^2\right)}{\left(1-\rho^2\right)^2} \\ &= \frac{1+\rho^2}{1-\rho^2} \end{aligned}$$

For the predictor that uses all three adjacent samples, the prediction coefficients $\mathbf{h}_3 = [h\ v\ d]^T$ are given by the solution of the linear equation system

$$\mathbf{C}_3\,\mathbf{h}_3 = \mathbf{c}_3$$

with

$$\mathbf{C}_3 = \sigma_S^2 \begin{bmatrix} 1 & \rho_D & \rho_V \\ \rho_D & 1 & \rho_H \\ \rho_V & \rho_H & 1 \end{bmatrix} = \sigma_S^2 \begin{bmatrix} 1 & \rho^2 & \rho \\ \rho^2 & 1 & \rho \\ \rho & \rho & 1 \end{bmatrix}$$

and

$$\mathbf{c}_3 = \sigma_S^2 \begin{bmatrix} \rho_H \\ \rho_V \\ \rho_D \end{bmatrix} = \sigma_S^2 \begin{bmatrix} \rho \\ \rho \\ \rho^2 \end{bmatrix}$$

Hence, we have

$$
\begin{aligned}
h + \rho^2 \cdot v + \rho \cdot d &= \rho \\
\rho^2 \cdot h + v + \rho \cdot d &= \rho \\
\rho \cdot h + \rho \cdot v + d &= \rho^2
\end{aligned}
$$

Multiplying the last equation by $-\rho$ and adding it to the first and second equation yields

$$
\begin{aligned}
(1 - \rho^2)\, h &= \rho\,(1 - \rho^2) \\
(1 - \rho^2)\, v &= \rho\,(1 - \rho^2)
\end{aligned}
$$

and, hence,

$$h = v = \rho$$

Inserting this in the third equation of the above equation system yields

$$d = -\rho^2$$

Hence, we have

$$\mathbf{h}_3 = \begin{bmatrix} \rho \\ \rho \\ -\rho^2 \end{bmatrix}$$

And for the MSE, we obtain

$$
\begin{aligned}
\varepsilon^2_{HVD} &= \sigma^2_S - \mathbf{c}_3^T\, \mathbf{h}_3 \\
&= \sigma^2_S - \sigma^2_S \begin{bmatrix} \rho & \rho & \rho^2 \end{bmatrix} \begin{bmatrix} \rho \\ \rho \\ -\rho^2 \end{bmatrix} \\
&= \sigma^2_S - \sigma^2_S \left(\rho^2 + \rho^2 - \rho^4 \right) \\
&= \sigma^2_S \left(1 - 2\rho^2 + \rho^4 \right) \\
&= \sigma^2_S \left(1 - \rho^2 \right)^2
\end{aligned}
$$

Hence, the prediction gain is

$$g_{HVD} = \frac{\sigma^2_S}{\varepsilon^2_{HVD}} = \frac{1}{\left(1 - \rho^2\right)^2}$$

For $\rho = 0.95$, we get the following prediction gains in dB:

$$
\begin{aligned}
G_H = 10\,\log_{10} \frac{1}{1 - \rho^2} &\approx 10.11 \text{ dB} \\
G_{HV} = 10\,\log_{10} \frac{1 + \rho^2}{1 - \rho^2} &\approx 12.90 \text{ dB} \\
G_{HVD} = 10\,\log_{10} \frac{1}{\left(1 - \rho^2\right)^2} &\approx 20.22 \text{ dB}
\end{aligned}
$$

23. 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.

<u>*Solution:*</u>

The covariance between two samples S_n and S_{n-1} is given by

$$\begin{aligned} E\{S_n S_{n-1}\} &= E\{(Z_n + \alpha_1 S_{n-1} + \alpha_2 S_{n-2})\, S_{n-1}\} \\ &= E\{Z_n\, S_{n-1}\} + \alpha_1\, E\{S_{n-1}^2\} + \alpha_2\, E\{S_{n-1}\, S_{n-2}\} \\ \rho_1\, \sigma_S^2 &= \alpha_1\, \sigma_S^2 + \alpha_2\, \rho_1\, \sigma_S^2 \\ \rho_1\, (1 - \alpha_2) &= \alpha_1 \end{aligned}$$

yielding

$$\rho_1 = \frac{\alpha_1}{1 - \alpha_2} = \frac{0.7}{1 - 0.2} = 0.875$$

Similarly, the covariance between two samples S_n and S_{n-2} is given by

$$\begin{aligned} E\{S_n S_{n-2}\} &= E\{(Z_n + \alpha_1 S_{n-1} + \alpha_2 S_{n-2})\, S_{n-2}\} \\ &= E\{Z_n\, S_{n-2}\} + \alpha_1\, E\{S_{n-1} S_{n-2}\} + \alpha_2\, E\{S_{n-2}^2\} \\ \rho_2\, \sigma_S^2 &= \alpha_1\, \rho_1\, \sigma_S^2 + \alpha_2\, \sigma_S^2 \\ \rho_2 &= \alpha_1\, \rho_1 + \alpha_2 \\ &= \alpha_1\, \frac{\alpha_1}{1 - \alpha_2} + \alpha_2 \\ &= \frac{\alpha_1^2 - \alpha_2^2 + \alpha_2}{1 - \alpha_2} \end{aligned}$$

yielding

$$\rho_2 = \frac{\alpha_1^2 - \alpha_2^2 + \alpha_2}{1 - \alpha_2} = \frac{0.7^2 - 0.2^2 + 0.2}{1 - 0.2} = 0.8125$$

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

Solution:
The predictor is given by

$$\hat{S_n} = [h_1 \; h_2] \begin{bmatrix} S_{n-1} \\ S_{n-2} \end{bmatrix}$$

The optimal predictor is the solution of the following linear equation system

$$\begin{aligned} \mathbf{C}_2 \cdot \mathbf{h} &= \mathbf{c}_1 \\ \sigma_S^2 \begin{bmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{bmatrix} \cdot \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} &= \sigma_S^2 \begin{bmatrix} \rho_1 \\ \rho_2 \end{bmatrix} \end{aligned}$$

or

$$\begin{aligned} h_1 + \rho_1 \, h_2 &= \rho_1 \\ \rho_1 \, h_1 + h_2 &= \rho_2 \end{aligned}$$

yielding

$$\begin{aligned} (1 - \rho_1^2) \, h_1 &= \rho_1 - \rho_1 \, \rho_2 \\ h_1 &= \rho_1 \, \frac{1 - \rho_2}{1 - \rho_1^2} \\ &= 0.875 \cdot \frac{1 - 0.8125}{1 - 0.875^2} \\ h_1 &= 0.7 \end{aligned}$$

and

$$\begin{aligned} h_2 &= \rho_2 - h_1 \, \rho_1 \\ &= 0.8125 - 0.7 \cdot 0.875 \\ h_2 &= 0.2 \end{aligned}$$

It should be noted that the prediction coefficients are equal to the AR parameters ($h_1 = \alpha_1$ and $h_2 = \alpha_2$), which can be shown by inserting the formulas for the correlation factors in the above equations.
For the variance of the prediction error, we obtain

$$\begin{aligned} \sigma_U^2 &= \sigma_S^2 - \mathbf{h}^T \cdot \mathbf{c}_1 \\ &= \sigma_S^2 - h_1 \cdot \rho_1 \sigma_S^2 - h_2 \cdot \rho_2 \sigma_S^2 \\ &= \sigma_S^2 \, (1 - h_1 \cdot \rho_1 - h_2 \cdot \rho_2) \\ &= \sigma_S^2 \, (1 - 0.7 \cdot 0.875 - 0.2 \cdot 0.8125) \\ \sigma_U^2 &= 0.225 \cdot \sigma_S^2 \end{aligned}$$

yielding the prediction gain

$$\begin{aligned} G_P &= 10 \cdot \log_{10} \frac{\sigma_S^2}{\sigma_U^2} \\ &= 10 \cdot \log_{10} \frac{1}{0.225} \\ G_P &\approx 6.4782 \, \text{dB} \end{aligned}$$

(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?

Solution:

The optimal prediction coefficient is the solution of the equation

$$\begin{aligned} \mathbf{C}_1 \cdot h &= \mathbf{c}_1 \\ \sigma_S^2 \cdot h &= \rho_1 \cdot \sigma_S^2 \end{aligned}$$

yielding

$$h = \rho_1 = 0.875$$

The resulting prediction error variance is given by

$$\begin{aligned} \sigma_U^2 &= \sigma_S^2 - h \cdot \mathbf{c}_1 \\ &= \sigma_S^2 - \rho_1 \cdot \rho_1 \cdot \sigma_S^2 \\ &= \sigma_S^2 \, (1 - \rho_1^2) \\ &= \sigma_S^2 \, (1 - 0.875^2) \\ \sigma_U^2 &= 0.234375 \cdot \sigma_S^2 \end{aligned}$$

yielding the prediction gain

$$\begin{aligned} G_P &= 10 \cdot \log_{10} \frac{\sigma_S^2}{\sigma_U^2} \\ &= 10 \cdot \log_{10} \frac{1}{0.234375} \\ G_P &\approx 6.3009 \, \mathrm{dB} \end{aligned}$$

The loss versus optimal prediction is

$$\begin{aligned} L &= 10 \cdot \log_{10} \frac{0.234375 \, \sigma_S^2}{0.225 \, \sigma_S^2} \\ &= 10 \cdot \log_{10} \frac{0.234375}{0.225} \\ &\approx 0.1773 \, \mathrm{dB} \end{aligned}$$

(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?

<u>*Solution:*</u>
The covariance of the prediction error after the first stage is given by

$$\begin{aligned}
E\{U_n\,U_{n-1}\} &= E\{(S_n - \rho_1\,S_{n-1})\,(S_{n-1} - \rho_1\,S_{n-2})\} \\
&= E\{S_n\,S_{n-1}\} - \rho_1\,E\{S_{n-1}^2\} - \rho_1\,E\{S_n\,S_{n-2}\} \\
&\quad +\rho_1^2\,E\{S_{n-1}\,S_{n-2}\} \\
\rho_U\,\sigma_U^2 &= \rho_1\,\sigma_S^2 - \rho_1\,\sigma_S^2 - \rho_1\rho_2\,\sigma_S^2 + \rho_1^3\,\sigma_S^2 \\
&= \sigma_S^2 \cdot \rho_1\,(\rho_1^2 - \rho_2)
\end{aligned}$$

Since, the covariance is not equal to zero (for $\rho_2 \neq \rho_1^2$), the prediction can be improved by adding a second prediction stage.
Inserting the formula for the prediction error variance after the first prediction stage, $\sigma_U^2 = \sigma_S^2\,(1-\rho_1^2)$, which has been derived above, we obtain

$$\begin{aligned}
\rho_U\,\sigma_U^2 &= \sigma_S^2 \cdot \rho_1\,(\rho_1^2 - \rho_2) \\
\rho_U\,\sigma_S^2\,(1-\rho_1^2) &= \sigma_S^2 \cdot \rho_1\,(\rho_1^2 - \rho_2) \\
\rho_U &= \rho_1\,\frac{\rho_1^2 - \rho_2}{1-\rho_1^2} \\
&= 0.875 \cdot \frac{0.875^2 - 0.8125}{1 - 0.875^2} \\
&= -0.175
\end{aligned}$$

Hence, the optimal second stage predictor is given by

$$h = \rho_U = -0.175$$

and the final prediction error becomes

$$\begin{aligned}
\sigma_V^2 &= \sigma_U^2\,(1-\rho_U^2) \\
&= 0.969375\,\sigma_U^2
\end{aligned}$$

yielding the second stage prediction gain

$$\begin{aligned}
G_p &= 10 \cdot \log_{10} \frac{\sigma_U^2}{\sigma_V^2} \\
&= 10 \cdot \log_{10} \frac{1}{0.969375} \\
&\approx 0.1351 \text{ dB}
\end{aligned}$$

For the loss versus optimal prediction using the last two samples, we obtain

$$\begin{aligned} L &= 10\log_{10}\frac{0.969375\,\sigma_U^2}{0.225\,\sigma_S^2} \\ &= 10\log_{10}\frac{0.969375\cdot 0.234375\,\sigma_S^2}{0.225\,\sigma_S^2} \\ &= 10\log_{10}\frac{0.969375\cdot 0.234375}{0.225} \\ &\approx 0.0422\text{ dB} \end{aligned}$$

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.

Solution:
The optimal linear predictor is given by

$$\hat{S}_N = h \cdot S_{n-1} \qquad \text{with} \qquad h = \rho$$

The resulting prediction error is

$$\sigma_U^2 = \sigma_S^2 \, (1 - \rho^2)$$

Hence, we have the prediction gain

$$\begin{aligned} G_P &= 10 \cdot \log_{10} \frac{\sigma_S^2}{\sigma_U^2} \\ &= 10 \cdot \log_{10} \frac{1}{1-\rho^2} \\ &= 10 \cdot \log_{10} \frac{1}{1-0.9^2} \\ &\approx 7.2125 \text{ dB} \end{aligned}$$

(b) Use the predictor derived in (24a) 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?

Solution:
For the prediction samples, we can write

$$\hat{S}_n = h \cdot S'_{n-1} = h \cdot (S_{n-1} - Q_{n-1}),$$

where $Q_n = S_n - S'_n$ represents the quantization error.
For the prediction error, we obtain

$$U_n = S_n - \hat{S}_n = S_n - h\, S_{n-1} + h\, Q_{n-1}$$

The mean of the prediction error is

$$\begin{aligned} \mu_U &= E\{U_n\} = E\{S_n\} - h\, E\{S_n\} + h\, E\{Q_n\} \\ &= \mu_S(1-h) + \mu_Q \\ &= 0 \end{aligned}$$

Note that $\{S_n\}$ is a zero-mean process and μ_Q is zero, since an entropy-constrained quantizer is a centroidal quantizer for which the mean of the quantization error is always zero.
For the prediction error variance, we obtain

$$\begin{aligned}
\sigma_U^2 &= E\{U_n^2\} = E\{(S_n - h\,S_{n-1} + h\,Q_{n-1})^2\} \\
&= E\{S_n^2\} - 2h\,E\{S_n\,S_{n-1}\} + 2h\,E\{S_n\,Q_{n-1}\} \\
&\quad + h^2\,E\{S_{n-1}^2\} + h^2\,E\{Q_{n-1}^2\} - 2h^2\,E\{S_{n-1}Q_{n-1}\} \\
&= \sigma_S^2 - 2h\rho\sigma_S^2 + h^2\sigma_S^2 + h^2\sigma_Q^2 \\
&\quad + 2h\,E\{S_n\,Q_{n-1}\} - 2h^2\,E\{S_{n-1}Q_{n-1}\} \\
&= \sigma_S^2 - 2\rho^2\sigma_S^2 + \rho^2\sigma_S^2 + \rho^2\sigma_Q^2 \\
&\quad + 2\rho\,E\{S_n\,Q_{n-1}\} - 2\rho^2\,E\{S_{n-1}Q_{n-1}\} \\
&= (1-\rho^2)\,\sigma_S^2 + \rho^2\sigma_Q^2 + 2\rho\,E\{S_n\,Q_{n-1}\} - 2\rho^2\,E\{S_{n-1}Q_{n-1}\}
\end{aligned}$$

For the cross-term, we have

$$\begin{aligned}
&2\rho\,E\{S_n\,Q_{n-1}\} - 2\rho^2\,E\{S_{n-1}Q_{n-1}\} \\
&\quad= 2\rho\,E\{(Z_n + \rho S_{n-1})Q_{n-1}\} - 2\rho^2\,E\{S_{n-1}Q_{n-1}\} \\
&\quad= 2\rho\,E\{Z_nQ_{n-1}\} + 2\rho^2\,E\{S_{n-1}Q_{n-1}\} - 2\rho^2\,E\{S_{n-1}Q_{n-1}\} \\
&\quad= 2\rho\,E\{Z_nQ_{n-1}\} \\
&\quad= 0
\end{aligned}$$

yielding the prediction error variance

$$\sigma_U^2 = (1-\rho^2)\,\sigma_S^2 + \rho^2\sigma_Q^2$$

For high rates (which we consider), the quantization error can be expressed using the high-rate approximation for the entropy-constrained Lloyd quantizer for Gaussian sources

$$\sigma_Q^2 = \frac{\pi e}{6}\cdot\sigma_U^2\cdot 2^{-2R}$$

yielding

$$\begin{aligned}
\sigma_U^2 &= (1-\rho^2)\,\sigma_S^2 + \rho^2\cdot\frac{\pi e}{6}\cdot\sigma_U^2\cdot 2^{-2R} \\
\sigma_U^2\left(1 - \frac{\pi e}{6}\cdot\rho^2\cdot 2^{-2R}\right) &= (1-\rho^2)\,\sigma_S^2 \\
\sigma_U^2 &= \sigma_S^2\cdot\frac{1-\rho^2}{1-\frac{\pi e}{6}\cdot\rho^2\cdot 2^{-2R}} \\
\sigma_U^2 &= 6\cdot\sigma_S^2\cdot\frac{1-\rho^2}{6-\pi\cdot e\cdot\rho^2\cdot 2^{-2R}}
\end{aligned}$$

And for the quantization error (for high rates), we get

$$\begin{aligned}
\sigma_Q^2 &= \frac{\pi e}{6}\cdot\sigma_U^2\cdot 2^{-2R} \\
&= \frac{\pi e}{6}\cdot\left(6\cdot\sigma_S^2\cdot\frac{1-\rho^2}{6-\pi\cdot e\cdot\rho^2\cdot 2^{-2R}}\right)\cdot 2^{-2R} \\
&= \sigma_S^2\cdot\left(\frac{\pi e(1-\rho^2)}{6-\pi\cdot e\cdot\rho^2\cdot 2^{-2R}}\right)\cdot 2^{-2R}
\end{aligned}$$

Without prediction, we would have the high-rate quantization error

$$\sigma_{Q0}^2 = \frac{\pi e}{6} \cdot \sigma_S^2 \cdot 2^{-2R}$$

Hence, the coding gain (for high rates) is given by

$$\begin{aligned} G_C(R) &= 10 \cdot \log_{10} \frac{\sigma_{Q0}^2}{\sigma_Q^2} \\ &= 10 \cdot \log_{10} \frac{6 - \pi e \rho^2 \cdot 2^{-2R}}{6 \cdot (1 - \rho^2)} \end{aligned}$$

The coding gain is dependent on the actual bit rate.
For the rate $R_1 = 1$, we obtain

$$\begin{aligned} G_C(R_1) &= 10 \cdot \log_{10} \frac{6 - \pi e \cdot 0.9^2 \cdot 2^{-2}}{6 \cdot (1 - 0.9^2)} \\ &\approx 5.7359 \text{ dB} \end{aligned}$$

For the rate $R_2 = 1$, we obtain

$$\begin{aligned} G_C(R_2) &= 10 \cdot \log_{10} \frac{6 - \pi e \cdot 0.9^2 \cdot 2^{-4}}{6 \cdot (1 - 0.9^2)} \\ &\approx 6.8877 \text{ dB} \end{aligned}$$

For the rate $R_3 = 3$, we obtain

$$\begin{aligned} G_C(R_3) &= 10 \cdot \log_{10} \frac{6 - \pi e \cdot 0.9^2 \cdot 2^{-6}}{6 \cdot (1 - 0.9^2)} \\ &\approx 7.1335 \text{ dB} \end{aligned}$$

For the rate $R_4 = 4$, we obtain

$$\begin{aligned} G_C(R_3) &= 10 \cdot \log_{10} \frac{6 - \pi e \cdot 0.9^2 \cdot 2^{-8}}{6 \cdot (1 - 0.9^2)} \\ &\approx 7.1929 \text{ dB} \end{aligned}$$

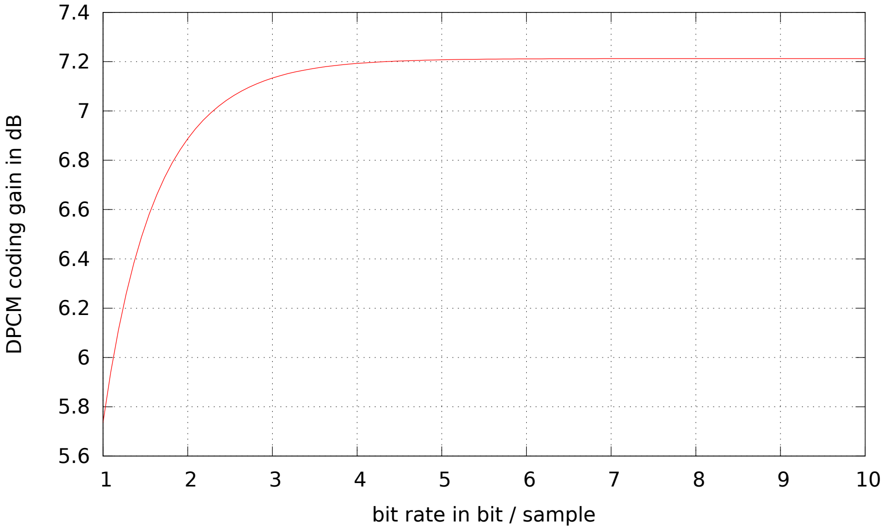
For the rate $R_5 = 8$, we obtain

$$\begin{aligned} G_C(R_3) &= 10 \cdot \log_{10} \frac{6 - \pi e \cdot 0.9^2 \cdot 2^{-16}}{6 \cdot (1 - 0.9^2)} \\ &\approx 7.2124 \text{ dB} \end{aligned}$$

The coding gain increase with increasing bit rate. For R appraoching infinity, the coding gain approaches the prediction gain:

$$\lim_{R \to \infty} G_C(R) = 10 \cdot \log_{10} \frac{1}{1 - 0.9^2} = G_P$$

In the following diagram, the DPCM coding gain is plotted as function of the bit rate. For bit rates less than 1 bit per sample, our high rate assumption is not valid.



DPCM coding gain in dB
7.4
7.2
7
6.8
6.6
6.4
6.2
6
5.8
5.6
1
2
3
4
5
6
7
8
9
10
bit rate in bit / sample